



# Analyse, modélisation, et détection automatique des disfluences dans le dialogue oral spontané contraint : le cas du Contrôle Aérien

Jean-Léon Mehdi Bouraoui

## ► To cite this version:

Jean-Léon Mehdi Bouraoui. Analyse, modélisation, et détection automatique des disfluences dans le dialogue oral spontané contraint : le cas du Contrôle Aérien. Autre [cs.OH]. Université Paul Sabatier - Toulouse III, 2008. Français. <tel-00354772>

**HAL Id: tel-00354772**

**<https://tel.archives-ouvertes.fr/tel-00354772>**

Submitted on 20 Jan 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse III - Paul Sabatier  
Discipline ou spécialité : Informatique

---

Présentée et soutenue par Jean-Léon Mehdi BOURAOUI  
Le 30 octobre 2008

**Titre :** *Analyse, modélisation, et détection automatique des disfluences dans le dialogue oral spontané contraint : le cas du Contrôle Aérien*

---

### JURY

Régine André-Obrecht, Co-directrice  
Nadine Vigouroux, Co-directrice  
Maryvonne Abraham, Rapporteur  
Jean Caelen, Rapporteur  
Jean-Luc Nespoulous, Examineur  
Robert Parisse, Examineur

---

**Ecole doctorale :** *Mathématiques Informatique Télécommunications de Toulouse*

**Unité de recherche :** UMR 5505

**Directeur(s) de Thèse :** Régine André-Obrecht, Nadine Vigouroux

**Rapporteurs :** Maryvonne Abraham, Jean Caelen







*A mes grand-parents,  
à mes parents,  
à mes frères et sœur*



# Remerciements

En premier lieu, je remercie évidemment Nadine Vigouroux à qui je dois tant. Non seulement pour son encadrement rigoureux, mais surtout pour son humanité et sa confiance. Je pourrais remplir le reste de ce document en louanges à son égard. En un mot comme en cent : MERCI. Mes remerciements vont également à Régine André-Obrecht, pour avoir co-encadré mon travail.

Je remercie Maryvonne Abraham et Jean Caelen pour avoir accepté d'être rapporteurs de mon travail, et pour avoir accepté de bonne grâce des conditions de temps très restreintes.

Toute ma gratitude va aussi à Jean-Luc Nespoulous d'avoir accepté d'être président de mon jury, ainsi que pour sa grande chaleur humaine, sa confiance et ses encouragements.

Je dois également beaucoup à Philippe Truillet. Que ce soit d'un point de vue scientifique et technique, ou pour son soutien psychologique. Merci aussi à Gwenael Bothorel pour avoir toujours été là lorsque j'avais des questions, et pour sa présence amicale. Toute ma reconnaissance également à Robert Parisse pour avoir accepté de faire partie de mon jury.

Merci en particulier à Igor, sans qui le module de détection ne serait pas ce qu'il est, pour son « coaching » durant des mois difficiles, et tout simplement pour sa générosité et sa patience.

J'ai évidemment une pensée très amicale pour tous ceux de l'ancienne équipe DIAMANT, que je n'ai pas encore cités jusqu'ici : Régis, Fabrice, Antonio, Mathieu, Bernard, Philippe, Florian, Mustapha, Christophe, Laurent, Guillaume, Caroline, Frédéric, Aurélie, Khaldoum, Youssef, ainsi que Jérôme et José, pour leur camaraderie. Avec une mention toute particulière pour Karine. Le déroulement de ma thèse, et ce qui l'a précédé, est loin d'avoir toujours été facile. Mais ne serait-ce que pour la rencontrer, cela valait la peine ! Je suis très heureux d'avoir la chance de bénéficier de son amitié.

Tous mes remerciements amicaux à celles et ceux qui se sont dévoués à la lecture attentive de tout ou partie de ma thèse : Karine, Laurent et Nathalie, Caroline, Florence, Jonathan, Guylhem, Philippe.

Mes remerciements s'adressent également à l'ensemble du personnel de l'IRIT, notamment Jean-Pierre, Jean-Claude, Agathe, Françoise, Cathy, Anne-Marie, Michèle, Cédric, Véronique, Chantal, Martine, ainsi que Martine Labruyère (de l'Ecole Doctorale) qui ont toujours été d'une gentillesse et d'une disponibilité à toute épreuve.



Je remercie aussi tous les membres du LIUPPA de Pau, pour leur accueil et la bonne ambiance qui y règne. Remerciements particuliers pour Jean-Michel Bruel et Alain Teste pour m'avoir aidé à concilier ma charge complète d'ATER et ma thèse, à mes collègues de bureau Damien Pallacio et Ten Lu, et à Régine.

Enfin, ces remerciements ne seraient pas complets sans l'expression de mon très profond amour pour mon père et ma mère, pour l'ensemble de ce que je leur dois, et leur affection. Merci mille fois.

Pour conclure, selon une formule consacrée dans les remerciements de thèse, je remercie celles et ceux qui n'auraient pas été nommés dans ces remerciements par oubli de ma part, et les prie de m'en excuser.

Finalement, les remerciements, c'est comme la thèse elle-même : on pourrait ne pas s'arrêter de les écrire !

# Table des matières

Table des matières .....	9
Liste des figures .....	14
Liste des tableaux.....	15
Introduction.....	17
 PARTIE 1 DEFINITIONS ET PROBLEMATIQUE .....	 23
Chapitre 1 Erreurs et disfluences dans le dialogue oral.....	25
1. Introduction.....	25
2. Le dialogue oral spontané .....	25
2.1. Principales caractéristiques de l'oral spontané .....	25
2.1.1. Linéarité .....	26
2.1.2. Souplesse de la syntaxe.....	27
2.1.3. Les structures clivées .....	27
2.1.4. Utilisation de la prosodie .....	27
2.1.5. Utilisation des interjections .....	27
2.1.6. Les disfluences.....	28
2.2. Principales caractéristiques des dialogues oraux .....	28
2.2.1. Les dialogues respectent des règles conventionnelles implicites .....	28
2.2.2. Les dialogues sont structurés .....	29
2.2.3. Les dialogues nécessitent une base commune .....	29
3. Les problèmes dans le dialogue .....	30
3.1. Erreurs et corrections .....	30
3.1.1. Théorie générale de l'erreur.....	31
3.1.2. Erreurs et corrections dans le dialogue oral.....	33
3.2. Les disfluences.....	35
3.2.1. Premières définitions .....	35
3.2.2. Blanche-Benveniste .....	37
3.2.3. Shriberg.....	38
3.2.3.1. Les répétitions ("repeated words").....	39
3.2.3.2. Les pauses remplies ("filled pauses") .....	39
3.2.3.3. Les termes explicites d'édition ("explicit editing terms") .....	40
3.2.3.4. Les marqueurs de discours ("discourse markers").....	40
3.2.3.5. Les fragments de mot ("word fragments").....	40
3.2.3.6. Insertions ("inserted word"), suppressions ("deleted word"), substitutions ("Word in substituted string" / "Substituted-string fragment") .....	40
3.2.4. Candea .....	41
3.2.4.1. Les pauses silencieuses .....	41
3.2.4.2. Les euh dits d'hésitation .....	41
3.2.4.3. Les allongements vocaliques .....	42
3.2.4.4. Les répétitions.....	42
3.2.4.5. Les autocorrections .....	43
3.2.4.6. Autres observations sur les disfluences .....	43

3.2.5.	Henry et Pallaud.....	44
3.2.5.1.	Les hésitations.....	45
3.2.5.2.	Les répétitions.....	45
3.2.6.	Kurdi .....	46
3.2.6.1.	Les extragrammaticalités lexicales .....	46
3.2.6.2.	Les extragrammaticalités supralexicales .....	47
4.	Conclusion .....	48
Chapitre 2 L'oral spontané dans le contrôle de la navigation aérienne .....		50
1.	Introduction.....	50
2.	Présentation générale du contexte.....	50
2.1.	Le trafic aérien .....	50
2.2.	La tâche de contrôle aérien .....	51
3.	Le dialogue oral dans le contrôle aérien .....	53
3.1.	Les dialogues contrôleurs aériens / pilotes .....	54
3.1.1.	Impératifs de la communication .....	54
3.1.2.	Caractéristiques de la phraséologie du contrôle aérien.....	54
3.2.	Analyses linguistiques des dialogues contrôleurs-pilotes.....	57
3.3.	Les dialogues d'apprentissage et d'évaluation .....	59
3.4.	La prévention des erreurs humaines dans le contrôle aérien .....	61
3.5.	Conséquences et caractéristiques cognitives et linguistiques .....	62
3.5.1.	Caractéristiques cognitives .....	62
3.5.2.	Caractéristiques linguistiques .....	64
4.	Corpus d'étude .....	66
4.1.	Composante orale .....	66
4.1.1.	Contexte d'acquisition .....	66
4.1.1.1.	Corpus A .....	66
4.1.1.2.	Corpus B .....	67
4.1.2.	Caractéristiques techniques.....	67
4.1.2.1.	Corpus A .....	67
4.1.2.2.	Corpus B .....	68
4.2.	Composante écrite.....	69
4.2.1.	Transcription et annotation .....	69
4.2.1.1.	Corpus A .....	69
4.2.1.2.	Corpus B .....	73
5.	Conclusion .....	74
PARTIE 2 ANALYSE ET MODELISATION.....		75
Chapitre 3 Analyse statistique et interprétative des phénomènes.....		77
1.	Introduction.....	77
2.	Méthodologie d'analyse des disfluences .....	78
2.1.	Méthode générale.....	78
2.2.	Typologie et terminologie.....	78
2.3.	Marqueurs recherchés .....	80
2.3.1.	Marqueurs lexicaux.....	80
2.3.2.	Marqueurs intonationnels .....	80
2.3.3.	Marqueurs de discours spontané.....	81
3.	Présentation générale des résultats (Corpus A) .....	81
3.1.	Distribution des phénomènes de disfluences .....	81

3.1.1.	Répétitions .....	82
3.1.2.	Hésitations .....	83
3.1.3.	Amorces et fragments de mots.....	83
3.1.4.	Allongements .....	84
3.1.5.	Pauses longues .....	84
3.2.	Distribution des phénomènes d’erreurs et de correction .....	84
3.2.1.	Acteurs du dialogue et catégories de corrections.....	85
3.2.2.	Manifestations de surface de l’erreur (niveau phonétique) .....	87
3.2.3.	Catégorisation des parties du discours affectées.....	88
3.3.	Fréquence et distribution des marqueurs .....	91
3.3.1.	Distribution globale des marqueurs .....	91
3.3.2.	Corrélation entre marqueurs et types d’erreurs et de correction.....	93
3.3.3.	Corrélation entre marqueurs et catégories de discours .....	94
4.	Présentation générale des résultats (Corpus B).....	94
4.1.	Typologie des phénomènes recensés dans ce corpus.....	94
4.2.	Résultats obtenus .....	95
4.2.1.	Fréquence des disfluences .....	95
4.2.2.	Fréquence des corrections.....	96
5.	Discussion de ces résultats.....	97
5.1.	Point de vue linguistique.....	97
5.2.	Point de vue cognitif et psycholinguistique.....	98
5.2.1.	Réflexions sur les niveaux de production auxquels interviennent les erreurs	98
5.2.2.	Stratégies en présence d’une erreur .....	100
5.3.	Point de vue informatique.....	100
6.	Conclusion .....	102
Chapitre 4 Modélisation.....		103
1.	Introduction.....	103
2.	Modélisation des disfluences .....	103
2.1.	Critères de choix du formalisme .....	104
2.2.	Le patron d’annotation de Bear et al.....	105
2.2.1.	Fondements théoriques .....	105
2.2.2.	Méthodologie d’annotation.....	106
2.3.	Ajouts et modifications apportés au patron de Bear et al. ....	108
3.	Description des patrons observés.....	110
3.1.	Aperçu général .....	111
3.2.	Patrons sans amorces .....	114
3.2.1.	Patrons simples .....	114
3.2.2.	Patrons comportant une ou plusieurs additions .....	116
3.2.3.	Patrons comportant une ou plusieurs omissions .....	117
3.2.4.	Autres patrons .....	118
3.2.5.	Patrons de répétitions.....	118
3.3.	Patrons avec amorces.....	119
3.3.1.	Patrons simples .....	119
3.3.2.	Patrons d’amorces comportant une ou plusieurs additions ou omissions 121	
4.	Discussion.....	122
5.	Conclusion .....	125

PARTIE 3 TRAITEMENT AUTOMATIQUE DES DISFLUENCES .....	127
Chapitre 5 Vers un traitement automatique des disfluences .....	129
1. Introduction.....	129
2. Approches existantes .....	129
2.1. Niveau du traitement du signal .....	130
2.2. Niveau de la compréhension de la parole .....	131
2.2.1. Approches statistiques et probabilistes .....	132
2.2.2. Approches syntaxico-sémantiques.....	135
2.2.3. Approches par patrons .....	137
3. Approche proposée .....	138
3.1. Adaptation au corpus et au contexte .....	138
3.2. Principaux éléments de notre proposition.....	138
3.2.1. Identification des traitements à appliquer .....	139
3.2.2. Méthode de recherche .....	142
4. Conclusion .....	143
Chapitre 6 Implémentation .....	144
1. Introduction.....	144
2. Spécifications de l'implémentation .....	144
3. Présentation du cadre d'implémentation.....	145
3.1. Historique et objectifs .....	145
3.2. Architecture .....	145
3.2.1. Gestion de la communication inter-agents.....	145
3.2.2. Gestion de la composante orale .....	146
3.2.3. Description des principaux agents .....	147
3.3. Capacités et limites de VOICE .....	151
4. Implémentation de notre modèle .....	152
4.1. Modifications au niveau de la grammaire.....	152
4.2. Conception et réalisation de l'agent.....	153
4.2.1. Principe général de l'algorithme .....	153
4.3. Intégration dans VOICE .....	157
4.3.1. Intégration dans l'architecture de VOICE .....	157
4.3.2. Modification de la grammaire de reconnaissance.....	159
4.3.3. Langage utilisé.....	161
5. Conclusion .....	162
Chapitre 7 Validation.....	164
1. Introduction.....	164
2. Méthodologie de validation .....	164
2.1. Paramètres de la validation .....	164
2.2. Constitution du corpus de validation .....	165
2.3. Apparat .....	168
2.3.1. Corpus.....	168
2.3.2. Dispositif employé.....	170
3. Résultats de la validation .....	171
3.1. Temps d'exécution.....	171
3.2. Performances du module .....	172
3.2.1. Aperçu général .....	172
3.2.2. Auto-corrections .....	174

3.2.3.	Répétitions .....	175
3.3.	Analyses complémentaires .....	176
3.3.1.	Cause des erreurs .....	176
3.3.2.	Comparaison avec d'autres études.....	178
4.	Conclusion .....	180
Conclusion .....		181
Bibliographie .....		189
Annexe 1 : Description de l'algorithme principal .....		203
Annexe 2 : Exemple d'application de l'algorithme .....		213
Annexe 3 : Exemples de transcriptions.....		215

## Liste des figures

Figure 1 : Axe syntagmatique et axe paradigmatique.....	26
Figure 2 : Présentation simplifiée des principaux niveaux du modèle de Garrett .....	34
Figure 3 : Modèle des disfluences de Shriberg .....	38
Figure 4 : Exemple d'application du modèle de Shriberg .....	39
Figure 5 : Contrôleurs aériens en exercice.....	52
Figure 6 : Exemple d'image radar .....	52
Figure 7 : Exemple d'un strip .....	52
Figure 8 : Communication pseudo-pilote / Contrôleur .....	61
Figure 9 : Hiérarchisation des sections et des tours de parole .....	73
Figure 10 : Distribution des disfluences dans le corpus A .....	82
Figure 11 : Espace de caractérisation d'une erreur de performance .....	85
Figure 12 : Distribution des corrections selon la fonction de la partie affectée .....	90
Figure 13 : Synthèse de la distribution des principaux patrons de correction sans amorce .....	122
Figure 14 : Synthèse de la distribution des principaux patrons d'amorces (correctives ou non) .....	122
Figure 15 : Rapport entre la taille des patrons et leur fréquence .....	124
Figure 16 : Un exemple d'auto-correction, avec représentation des différentes zones (tiré de Johnson <i>et al.</i> (2004b, p. 34)).....	132
Figure 17 : Représentation graphique des dépendances entre les zones de l'auto-correction de la Figure 16 (tiré de Johnson <i>et al.</i> ( <i>op. cit.</i> )).....	133
Figure 18 : Représentation syntaxique d'un énoncé disfluent avant traitement spécifique (système TiLT) .....	136
Figure 19 : Représentation syntaxique d'un énoncé disfluent après traitement spécifique (système TiLT) .....	136
Figure 20 : Gestion des connaissances linguistiques et contextuelles dans VOICE. ....	148
Figure 21 : Architecture de VOICE.....	150
Figure 22 : Intégration dans VOICE avec grammaire développée .....	157
Figure 23 : Intégration dans VOICE avec grammaire générique .....	158
Figure 24 : Rapport entre le nombre de mots et le temps d'exécution (échantillon 1).....	172
Figure 25 : Rapport entre le nombre de mots et le temps d'exécution (échantillon 2).....	172
Figure 26 : Taux de catégorisation correcte des disfluences par notre module.....	174

## Liste des tableaux

Tableau 1 : Présentation synoptique des catégorisations de disfluences .....	49
Tableau 2 : principales caractéristiques du corpus A .....	68
Tableau 3 : principales caractéristiques du corpus B.....	69
Tableau 4 : Comparaisons pour les répétitions .....	83
Tableau 5 : Comparaisons pour les hésitations.....	83
Tableau 6 : Comparaisons pour les allongements .....	84
Tableau 7 : Comparaisons pour les pauses longues.....	84
Tableau 8 : Nombre et pourcentage des stratégies de correction .....	86
Tableau 9 : Distribution des manifestations de surface (nombre et pourcentage).....	88
Tableau 10 : Distribution des différents types d'erreurs touchant un mot (nombre et pourcentage).....	91
Tableau 11 : Nombre et pourcentage des marqueurs (toutes catégories confondues).....	92
Tableau 12 : Nombre et pourcentage des marqueurs lexicaux .....	92
Tableau 13 : Distribution des différents marqueurs en fonction des catégories de correction (nombre d'occurrences et pourcentage).....	93
Tableau 14 : Distribution des marqueurs lexicaux en fonction des catégories de correction (nombre d'occurrences et pourcentage).....	93
Tableau 15 : Distribution des différents marqueurs en fonction des catégories de discours (nombre d'occurrences et pourcentage).....	94
Tableau 16 : Distribution des phénomènes de disfluences dans le corpus B (nombre d'occurrences et pourcentage) .....	96
Tableau 17 : Distribution des marqueurs de correction dans le corpus B (nombre d'occurrences et pourcentage) .....	96
Tableau 18 : Adaptation du patron de Bear <i>et al.</i> à notre corpus .....	109
Tableau 19 : Distribution des principaux types de patrons d'auto-correction selon les catégories fonctionnelles.....	113
Tableau 20 : Composition et distribution des patrons simples (sans amorce).....	114
Tableau 21 : Composition et distribution des patrons comportant au moins une addition (sans amorce) .....	116
Tableau 22 : Composition et distribution des patrons comportant au moins une omission (sans amorce) .....	118
Tableau 23 : Composition et distribution des autres catégories de patrons sans amorce ..	118
Tableau 24 : Composition et distribution des patrons de répétitions disfluentes .....	119
Tableau 25 : Composition et distribution des patrons simples avec amorce non-corrective .....	120
Tableau 26 : Composition et distribution des patrons simples avec amorce corrective ....	120
Tableau 27 : Composition et distribution des patrons d'addition ou omission avec amorce corrective .....	121
Tableau 28 : Comportements de recherche à adopter selon le type de disfluences.....	142
Tableau 29 : Caractéristiques techniques de la grammaire initiale de VOICE .....	159
Tableau 30 : Caractéristiques techniques de la grammaire modifiée de VOICE .....	160
Tableau 31 : Caractéristiques du corpus de test.....	170



Tableau 32 : Distribution des disfluences dans le corpus de test.....	170
Tableau 33 : Temps d'exécution (échantillon 1) .....	171
Tableau 34 : Temps d'exécution (échantillon 2) .....	171
Tableau 35 : Performances globales .....	173
Tableau 36 : Performances de traitement des auto-corrections .....	175
Tableau 37 : Performances de traitement des répétitions .....	175
Tableau 38 : Pourcentage des différentes causes d'erreurs de traitement .....	177
Tableau 39 : Présentation synoptique des performances d'autres systèmes de détection/traitement des disfluences .....	179

## Introduction

La modalité orale est la forme de communication la plus courante, et la plus ancienne. Son utilisation présente plusieurs avantages par rapport à l'écrit. Les plus souvent citées sont la rapidité et la convivialité. Il n'est pas étonnant que cette modalité fasse l'objet de recherches en Traitement Automatique du Langage et en Traitement Automatique de la Parole. Les applications sont en effet multiples : dialogue oral homme-machine, extraction d'information à partir d'enregistrements audio, pour ne citer que les plus répandues.

Cependant traiter automatiquement des productions orales spontanées pose de nombreux problèmes. Certains sont inhérents à toute communication, quelle que soit sa modalité : écrite ou orale. Utiliser un langage implique de prendre en compte des paramètres difficilement modélisables et implémentables. Ces paramètres sont, par exemple, les utilisations à des références ambiguës (anaphores, déictiques), et plus généralement la liberté offerte par les règles syntaxiques de ce langage.

A ces problèmes, rencontrés à l'écrit, s'en ajoutent d'autres si l'on considère spécifiquement l'oral. En effet, les spécificités de l'oral spontané rendent son traitement automatique encore plus délicat. Celles-ci sont appelés *disfluences*. Les disfluences affectent la production – et éventuellement la compréhension – de la production orale. La zone touchée de la production peut être locale (insertion d'hésitations par exemple) ou plus large (perturbation de la syntaxe).

En termes de traitement automatique, les disfluences posent des problèmes à différents niveaux. Au niveau du traitement du signal audio, il a été montré qu'elles engendrent des erreurs supplémentaires de reconnaissance des mots (Adda-Decker *et al.* (2004), p. 4). Au niveau de la compréhension automatique, elles peuvent entraîner des problèmes d'interprétation du sens, surtout lorsque la structure de l'énoncé a été modifiée (Bousquet (2002), Goulian (2002), Kurdi (2003)).

L'attitude la plus répandue face aux disfluences consiste à les ignorer lors de la reconnaissance du message oral. Autrement dit, leurs occurrences sont effacées, ou catégorisées comme « poubelle ». Dans la plupart des cas, cette approche est suffisante compte tenu de la robustesse des systèmes de reconnaissance de la parole. Autrement dit, elle n'entame pas la validité du traitement effectué par ailleurs.

Cependant, étudier et prendre en compte les disfluences, au lieu de les ignorer, présente deux principaux intérêts. Sur le plan théorique, mieux comprendre et modéliser les problèmes pouvant advenir lors de toute communication orale. Sur le plan applicatif également, les intérêts sont nombreux : par exemple, améliorer la robustesse des systèmes automatiques de dialogue oral (Antoine *et al.*, 2003), ou « nettoyer » automatiquement des transcriptions de productions orales spontanées (Adda-Decker, *op. cit.*).

Ces intérêts sont encore plus manifestes lorsque l'on s'intéresse à des domaines où des problèmes de communication peuvent être critiques. En effet, la majorité des études sur les disfluences portent, à des degrés divers, sur un langage non contraint : dialogues « à bâtons rompus », demandes d'horaire, discours, etc. Dans ce type de communication, les disfluences posent peu de problèmes, en raison notamment du peu d'enjeu qu'elles impliquent, et de la tolérance aux disfluences.

Mais il existe des domaines de l'activité humaine particulièrement critiques, pour lesquels la communication doit être autant que possible dépourvue de tout défaut. Le trafic aérien, et plus particulièrement son contrôle, en fait partie. Les communications verbales efficaces constituent un élément essentiel de sécurité dans le système aéronautique. L'utilisation appropriée des principes de communication devient encore plus importante quand les services de la circulation aérienne traitent avec des transporteurs aériens étrangers, dont le personnel navigant ne maîtrise pas toujours la langue anglaise. Le débit convenable de parole, le ton et la formulation, l'utilisation de la terminologie et des procédures de contrôle standard, contribuent à réduire le risque de malentendu entre les contrôleurs et les équipages. Les dialogues de contrôle de trafic aérien sont donc très contraints, à la fois d'un point de vue pragmatique (situation dans laquelle se trouvent les locuteurs) et linguistique (utilisation d'une phraséologie gouvernant ce qui doit être dit).

Dans ce contexte, les disfluences peuvent poser des problèmes critiques. Or, aucune étude n'a, à notre connaissance, été menée sur les disfluences apparaissant dans le contexte du contrôle aérien. Plus généralement, nous n'avons pas connaissance de travaux consacrés aux disfluences dans un contexte contraint. Or, nous faisons l'hypothèse que les disfluences présentent des particularités dans ce type de situation, différentes de celles mises à jour par les études existantes. Cette hypothèse est basée sur les observations, rapportées dans la littérature (par exemple Pierrel (2000)), des propriétés des langages contraints, ainsi que des caractéristiques des disfluences elles-mêmes.

Notre travail de thèse est consacré à la vérification de cette hypothèse. Plus précisément, nous poursuivons deux grands objectifs qui en découlent.

**Le premier de ces objectifs** est une étude exhaustive des phénomènes de disfluences dans le contexte particulier du dialogue de contrôle aérien. Ce sera notre principale manière d'étudier non seulement l'analyse de ces phénomènes, mais également leur modélisation.

**Notre deuxième objectif** est la proposition d'une méthodologie de traitement automatique adaptée à ce contexte. Il existe en effet des méthodologies efficaces de traitement automatique des disfluences (notamment Goulian (2002), Kurdi (2003)). Cependant, elles sont adaptées à des productions orales spontanées faites dans un langage non contraint. Or, si les disfluences présentent un comportement différent du fait de la phraséologie, cela implique également des modifications par rapport à leur traitement automatique. Il serait par conséquent intéressant, dans ce cas, de proposer une méthode adaptée.

On pourrait toutefois nous objecter que dans le contexte du contrôle aérien, un traitement automatique des disfluences est inutile. En effet, les interactions orales s'y font exclusivement entre humains. Pourquoi alors un traitement automatique des disfluences, puisque aucun système de compréhension automatique n'intervient lors de ces interactions ? Nous présentons dans ce document plusieurs applications pour lesquelles ce type de traitement est nécessaire. Nous évoquons notamment, dans le chapitre 2, le cas de l'apprentissage des contrôleurs aériens. Une partie de l'entraînement des contrôleurs consiste à interagir avec des personnes jouant le rôle de pilotes en vol ; le but étant de simuler aussi réalistement que possible le contrôle aérien en situation réelle. Dans ce cadre, il est possible, et même envisagé, que plusieurs des avions simulés soient pris en charge par un dispositif automatique. Le rôle de celui-ci est de comprendre les ordres donnés par le contrôleur en formation, et de faire évoluer les avions simulés selon ces ordres. On retrouve ici la problématique de la robustesse de la compréhension automatique de la parole, et de la détection automatique des disfluences afin d'améliorer cette robustesse. D'autre part, nous avons déjà mentionné le caractère critique du contrôle aérien : toute erreur ou incompréhension peut entraîner des conséquences dramatiques. De ce fait, des études prospectives <sup>1</sup> envisagent la mise en place d'un dispositif permettant de signaler aux contrôleurs et/ou aux pilotes l'occurrence de problèmes de compréhension (par exemple, différence entre un ordre donné par le contrôleur et la confirmation qui en est faite par le pilote). Ce type de dispositif est basé sur un système de compréhension automatique de la parole ; de ce fait, lui aussi bénéficierait de la robustesse apportée par la reconnaissance des disfluences. D'autant plus que celles-ci peuvent engendrer des problèmes de compréhension ; dans ce contexte applicatif précis, leur détection serait également particulièrement intéressante.

---

<sup>1</sup> Cf Deherty *et al.* 2005, particulièrement p. 6-7.

D'une manière plus générale, prendre en compte les disfluences dans toute interaction médiée relative au contrôle aérien, c'est contribuer à l'amélioration de la sécurité de ce type de communication, comme nous le montrerons également.

Ce rapport de thèse présente le travail que nous avons entrepris afin d'atteindre ces objectifs. Il est composé de trois grandes parties, elles mêmes composées de chapitres.

La première partie a trois objectifs, dont le point commun est la définition de notre problématique et de son contexte. D'une part, circonscrire notre espace de recherche, c'est-à-dire déterminer quels sont les aspects précis que nous traiterons. D'autre part, donner les définitions qui seront utilisées dans la suite du travail. Enfin, présenter le contexte d'étude : le contrôle de trafic aérien, ainsi que les corpus que nous utilisons.

Dans un **premier chapitre**, nous nous livrons à un état de l'art des principales études portant sur les disfluences. Notre démarche est double : d'une part, établir comment ces phénomènes ont déjà été traités par d'autres chercheurs de différentes disciplines : sciences du langage, informatique, psycholinguistique, psychologie cognitive. D'autre part, pouvoir disposer d'une typologie, ou au moins de propositions de typologie des disfluences.

Le **second chapitre** est consacré à positionner notre étude dans le domaine d'application dans lequel elle se situe. En effet, nous montrons que la tâche de contrôle aérien présente des caractéristiques qui la rendent très spécifique, surtout en regard d'autres types de dialogue plus proches de la vie quotidienne.

Ce chapitre comporte également la description des corpus qui nous ont servi de base de travail. Il s'agit de deux corpus présentant une composante aussi bien orale qu'écrite. Orale, parce qu'ils résultent de l'enregistrement de dialogues oraux spontanés, mais écrite également puisque ces enregistrements ont fait l'objet de transcriptions et d'annotations. Ces corpus ont pour point commun d'être relatifs à la tâche de contrôle aérien, mais présentent également un certain nombre de différences. La principale est le contexte d'enregistrement : l'un fut enregistré dans le cadre de l'apprentissage de la phraséologie par les contrôleurs, alors que l'autre se situe dans un contexte de travail « réel ».

La deuxième partie de la thèse est consacrée à l'analyse que nous avons effectuée des disfluences dans notre corpus d'étude. Nous présentons de manière exhaustive et argumentée la méthodologie que nous avons adoptée, les résultats obtenus, et les conclusions que nous en tirons.

Dans le **troisième chapitre**, nous présentons l'analyse lexico-syntaxique que nous avons menée sur les corpus décrits dans le chapitre précédent. Dans un premier temps, nous exposons la méthodologie de recherche que nous avons mise en place, ainsi que la typologie que nous avons dressée. Nous donnons ensuite les résultats observés, en termes de distribution des phénomènes et de modes de manifestations ; pour chaque type de disfluency, nous comparons ces données avec les résultats obtenus par les travaux présentés dans le chapitre 1.

Le **chapitre 4** est consacré à la modélisation des disfluencies de notre corpus. Nous effectuons d'abord une présentation générale des principes qui sous-tendent le schéma en question, et la manière dont nous l'avons adapté à nos corpus. Ensuite, nous étudions la distribution des différents patrons observés dans notre corpus. Nous discutons ces résultats, et envisageons comment ils peuvent être utilisés dans le cadre d'un traitement informatisé des disfluencies.

Enfin la troisième et dernière partie présente le traitement automatique des disfluencies que nous proposons. Nous y expliquons en détail la méthode que nous proposons, et nos motivations. Nous montrons ainsi que nos choix sont dictés par les résultats obtenus dans la partie 2, ainsi que par le contexte de la tâche.

Dans le **cinquième chapitre**, nous discutons la méthodologie à utiliser pour un traitement informatique des disfluencies, dans le cadre des dialogues de contrôle de trafic aérien. Dans un premier temps, nous passons en revue, et classifions, les différentes méthodes utilisées dans la littérature.

Dans un deuxième temps, nous nous positionnons par rapport à ces différentes approches. Suite au développement de cette analyse, nous présentons les grandes lignes du traitement automatique que nous proposons, en tenant compte des contraintes d'implémentation de l'application visée.

Le **chapitre 6** est consacré à la présentation détaillée de l'implémentation que nous proposons de la méthode proposée dans le chapitre précédent. Dans un premier temps, nous décrivons le système dans lequel est inséré notre module. Il s'agit du système VOICE, qui permet le traitement automatique, et en temps réel, d'ordres simples donnés dans une tâche de contrôle aérien à l'oral.

Nous développons ensuite de manière approfondie l'algorithme que nous avons mis au point, le langage informatique utilisé pour l'implémentation, ainsi que la manière dont nous l'avons intégré dans VOICE.

Enfin, le **chapitre 7** présente le processus de validation du module que nous avons implémenté. Nous y montrons la méthodologie que nous avons définie, les résultats obtenus, et les commentaires que ceux-ci appellent.

# **PARTIE 1**

## **DEFINITIONS ET PROBLEMATIQUE**





# **Chapitre 1**

## **Erreurs et disfluences dans le dialogue oral**

### **1. Introduction**

Conformément à la démarche que nous avons adoptée pour l'ensemble de cette thèse, le rôle de ce premier chapitre est de commencer à délimiter la zone de recherche que nous allons explorer.

Ce chapitre est composé de deux parties. La première est consacrée à la présentation du cadre dans lequel apparaissent les disfluences : la production orale spontanée, et plus particulièrement la situation de dialogue ; nous verrons que certaines caractéristiques propres au dialogue oral spontané sont invariables, quel que soit le type de dialogue envisagé. Dans un deuxième temps, nous nous intéressons aux différents problèmes pouvant advenir dans ce cadre, parmi lesquels les disfluences. Nous envisageons cette question selon le double angle des causes et des manifestations des disfluences.

### **2. Le dialogue oral spontané**

L'emploi de l'expression de « dialogue oral spontané » peut laisser à penser qu'il s'agit d'un seul phénomène. Or, au vu des nombreuses études parues, il est nécessaire de considérer deux parties différentes. D'une part l'aspect oral spontané, et d'autre part l'aspect dialogue. En effet, chacun présente des caractéristiques qui lui sont propres, et qui s'additionnent lorsqu'on les considère ensemble.

Les deux prochaines sous-sections sont consacrées respectivement à chacun de ces deux aspects.

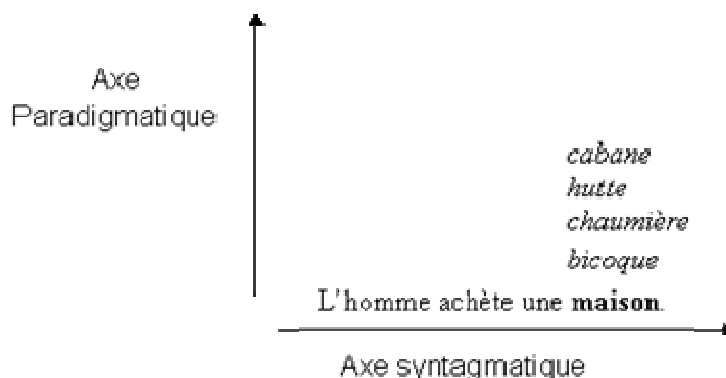
#### ***2.1. Principales caractéristiques de l'oral spontané***

Le terme de « spontané » est utilisé dans la littérature pour indiquer toute production orale qui n'a pas fait l'objet, en amont, d'une préparation rigoureuse (par exemple, discours appris par cœur... ou texte lu sur un prompteur !). Comme le note Blanche-Benveniste (1997, p. 17), ce type de productions est le plus courant.

Or ce caractère spontané de l'oral est à l'origine de la plupart de ses spécificités par rapport à l'écrit. Celles-ci sont bien connues, et ont abondamment été décrites dans la littérature. Nous nous baserons notamment sur les travaux de Blanche-Benveniste (1997) et Kerbrat-Orecchioni (2005) pour la liste suivante, non exhaustive.

### 2.1.1. Linéarité

L'adage le dit bien : « Verba volant, Scriba manent » (« les paroles s'envolent, les écrits restent »)<sup>2</sup>. Une fois qu'un mot a été prononcé, il n'est plus possible de revenir en l'arrière pour le corriger, ou ajouter ou retirer des éléments dans son contexte, contrairement à ce qui est possible avec l'écrit. Cette particularité de l'oral entraîne des conséquences sur la gestion par le locuteur de sa production. Par exemple, toute tentative de correction d'un élément ultérieur donne lieu à une perturbation de la structure attendue de l'énoncé. De même, le locuteur peut avoir à faire des pauses pour planifier la suite de son énonciation. Pour mieux comprendre cette notion et ce qu'elle implique, recourons au concept d'axes *syntagmatique* et *paradigmatique*. Ce concept a été initialement conçu par F. de Saussure dans ses *Cours de Linguistique Générale*<sup>3</sup> ; il est depuis fréquemment utilisé en linguistique, et s'applique à toute production, écrite ou orale. Pour résumer, l'axe syntagmatique est celui sur lequel s'enchaînent les différentes unités produites par le scripteur (ou locuteur) au fur et à mesure de sa production. Il peut être représenté par un axe horizontal. L'exemple typique en est l'énoncé. L'axe paradigmatique, au contraire, est à comprendre comme l'ensemble des choix possibles pour le scripteur (ou locuteur) à un moment donné de son énonciation. Par exemple<sup>4</sup>, soit l'énoncé « l'homme habite *une maison* » : le locuteur avait le choix de produire à la place du syntagme mis en italique : « une cabane », « une hutte », etc. Le choix dépend notamment des éléments précédents la position courante.



**Figure 1** : Axe syntagmatique et axe paradigmatique

<sup>2</sup> Nous nous en permettons un léger détournement : il ne s'applique pas au contexte que nous étudions ici !

<sup>3</sup> L'ouvrage publié n'est « que » le recueil par ses élèves des cours de Saussure ; il ne l'a pas rédigé lui-même

<sup>4</sup> Exemple inspiré de Blanche-Benveniste (1997, p. 16).

La Figure 1 ci-dessus illustre la différence entre ces deux axes. Nous montrerons en section 3.2.2 comment C. Blanche-Benveniste utilise ce concept pour rendre compte des disfluences.

### 2.1.2. *Souplesse de la syntaxe*

L'écrit est beaucoup plus restrictif sur l'utilisation de la syntaxe d'une langue que ne l'est l'oral. Ainsi, des constructions qui sont considérées comme des fautes à l'écrit sont tolérées à l'oral. Parmi les exemples les plus fréquents, on peut citer l'oubli de la négation « ne » (95% d'omission selon Blanche-Benveniste (*op. cit.*, p. 39)), l'interrogation indirecte, l'utilisation de l'interrogation « est-ce que », etc.

### 2.1.3. *Les structures clivées*

Voici la définition que donne Blanche-Benveniste de ces structures :

**« on appelle clivages, depuis les années 1980, les divers dispositifs par lesquels un élément est séparé et distingué du reste de sa construction »**  
**(Blanche-Benveniste (*op.cit.* p. 96)).**

Les structures clivées se présentent le plus souvent sous la forme « C'est X que Y ». Par exemple « c'est l'Europe qui négocie » (cité par Blanche-Benveniste, *op. cit.*, souligné dans le texte original). La mise en valeur d'un élément de l'énoncé peut également être effectuée par le moyen d'un *pseudo-clivage*, dont la forme la plus courante est « ce que X, c'est Y » (exemple : « ce que je voudrais, c'est recommencer à travailler », cité par Blanche-Benveniste (*op. cit.*, p. 98, souligné dans le texte original)) ;

### 2.1.4. *Utilisation de la prosodie*

La *prosodie* est un terme générique désignant la variation de paramètres acoustiques<sup>5</sup> de la voix. Elle a de multiples utilisations, dont certaines peuvent influencer sur la structure de l'énoncé. Citons à titre d'exemple la possibilité de marquer l'interrogation uniquement avec l'intonation montante. De ce fait, le locuteur peut éviter de recourir aux dispositifs utilisés à cette fin à l'écrit (comme par exemple l'inversion sujet verbe : « il est venu ? » VS « Est-il venu ? »).

### 2.1.5. *Utilisation des interjections*

Les *interjections* sont des mots utilisés pour exprimer les émotions du locuteur (exemple : « super ! »), ou reproduire des sons (on parle alors d'*onomatopées*). Elles peuvent également être employées à l'écrit, mais c'est la plupart du temps pour retranscrire des propos tenus à l'oral, ou simuler celui-ci.

---

<sup>5</sup> Hauteur, intensité, durée et timbre. Nous ne rentrons pas plus dans les détails car ce n'est pas l'objet principal de cette thèse. Le lecteur intéressé peut consulter par exemple Lacheret *et al.* (1999).

### 2.1.6. Les disfluences

Les *disfluences* sont, en apparence du moins, des perturbations de la production orale, telles que les hésitations. Elles font l'objet de cette thèse ; nous les présentons en détail en 3.2.

Enfin, la réception de la production orale peut subir différentes perturbations selon la qualité du canal et du contexte : environnement bruité, « parasites » sur une ligne téléphonique ...

### 2.2. Principales caractéristiques des dialogues oraux

Jusqu'ici, nous n'avons envisagé que les spécificités propres au discours oral spontané, c'est-à-dire une production d'un seul locuteur. Mais une situation courante est l'utilisation de l'oral pour communiquer avec autrui de manière interactive, ce que l'on désigne par le terme de dialogue. De fait, la notion de dialogue est bien plus subtile qu'il n'y paraît intuitivement. En effet, elle peut varier selon le type du dialogue, le contexte, la position des intervenants les uns par rapport aux autres (hiérarchie par exemple), etc.

Dans le cadre de cette thèse, nous considérerons qu'il y a dialogue dès lors qu'une communication orale spontanée s'établit entre aux moins deux agents humains<sup>6</sup>, pour remplir un but commun (même si ce but est la poursuite du dialogue lui-même, comme par exemple un dialogue au sujet de « la pluie et le beau temps »).

Mais quelle que soit la définition adoptée, les dialogues oraux spontanés présentent eux aussi un ensemble de caractéristiques qui leur sont propres. Nous en énumérons les principales ci-dessous ; une fois encore cette liste n'est pas exhaustive.

#### 2.2.1. Les dialogues respectent des règles conventionnelles implicites

H. P. Grice (1975) a été le premier à exposer ces règles. Pour cet auteur, toute conversation est dirigée par ce qu'il appelle le *principe de coopération*. Ce principe est formulé comme suit :

**« Que votre contribution à la conversation soit, au moment où elle intervient, telle que le requiert l'objectif ou la direction acceptée de l'échange verbal dans lequel vous êtes engagés »<sup>7</sup>**

Autrement dit, chaque locuteur est censé réagir de la manière la plus appropriée en fonction de l'état du dialogue (par exemple, il ne doit pas répondre à une question par une

---

<sup>6</sup> Dans le cadre de ce chapitre, nous n'envisageons pas la possibilité pour un humain de dialoguer avec un système informatique. Nous serions sinon contraints d'aborder une nouvelle problématique, qui trouvera partiellement sa place dans le chapitre 5 de ce rapport. Mais nous pouvons dès à présent noter qu'il est admis de parler de « dialogue » pour désigner la communication orale homme-machine.

<sup>7</sup> Cité dans Moeschler *et al.* (1994, p. 204).

autre question, passer subitement à un autre sujet, ...). Grice a formalisé de manière explicite les principes (appelés *maximes de Grice*) auxquels un locuteur devrait se conformer pour que le principe de coopération soit respecté. Chacune de ces maximes est une catégorie comprenant une ou plusieurs règles à suivre : de quantité (ne fournir que ce qui est nécessaire, mais en intégralité), de qualité (ne rien affirmer qui ne soit vrai et certain), de relation (être pertinent) ou de manière (être bref et ordonné).

L'une des principales caractéristiques des maximes de conversation est qu'elles permettent le déclenchement de ce que Grice nomme des *implicatures conventionnelles* : il s'agit d'inférences faites par un intervenant d'un dialogue sur le contenu réel du message de l'interlocuteur, compte tenu de la supposition que le principe de coopération (et donc les maximes) est respecté. Par exemple<sup>8</sup>, si un locuteur A demande à B où trouver de l'essence, et que B répond « Il y a un garage au bout de la rue », on considérera que si B respecte le principe de coopération (ici, la maxime de relation), l'énoncé implicite à la réponse de B est « Le garage est ouvert et on y trouve de l'essence ».

### 2.2.2. Les dialogues sont structurés

Les différents échanges des dialoguants (les participants au dialogue) sont structurés selon divers niveaux imbriqués les uns dans les autres. Cette modélisation du dialogue est partagée par plusieurs auteurs. Les plus célèbres sont ceux appartenant à ce que l'on a appelé « l'école de Genève » (voir par exemple Moeschler (1989), ou Roulet *et al.* (1985)). La désignation et la nature des niveaux modélisés varient selon les auteurs ; en général, les principaux sont, du plus bas au plus haut : les *actes de langage* (la plus petite unité), l'*intervention* (faite par un locuteur donné, et pouvant être constituée de plusieurs actes de langage), et l'*échange* (ensemble d'interventions relatives à un thème donné).

### 2.2.3. Les dialogues nécessitent une base commune

La manière dont les différents participants d'un dialogue mettent en commun les données qui leur seront nécessaires pour se comprendre est une problématique centrale de la pragmatique. D. Traum (voir notamment Traum (1999)), emploie le terme de *grounding*. Ce processus consiste, au début du dialogue, à acquérir et partager les connaissances communes ; dans la suite du dialogue, le *grounding* désignera la mise à jour de ces connaissances au fur et à mesure de l'évolution de la situation dans laquelle se situent les dialoguants. Cette modélisation du dialogue est particulièrement utilisée dans l'étude des dialogues collaboratifs (les dialoguants ont une tâche commune à accomplir, et utilisent le dialogue pour travailler en commun dans ce but).

---

<sup>8</sup> Moeschler *et al.* (*op. cit.*, p.206).

### 3. Les problèmes dans le dialogue

Jusqu'à présent, nous avons présenté les principales caractéristiques de la modalité des dialogues oraux spontanés. Parmi ces caractéristiques, nous avons montré que figurent certaines causes de problèmes potentiels, tels que les problèmes dus au canal de communication, les disfluences, ou encore l'incompréhension dans un dialogue due par exemple au non respect de l'une des règles citées en 2.2.

Nous allons maintenant examiner plus en détails la question de ces problèmes. Pour cela, et uniquement à des fins didactiques, nous les distribuons en deux groupes : d'une part les erreurs et corrections, et d'autre part les disfluences. Cette distinction que nous établissons entre les erreurs et corrections d'un côté, et les disfluences de l'autre, ne correspond pas à une prise de position ontologique. En effet, les corrections sont souvent considérées comme des disfluences, comme nous le verrons ; et à l'inverse, il est fréquent que les disfluences apparaissent au niveau des erreurs ou des corrections.

Nous maintenons cependant cette distinction pour deux raisons. D'une part, les deux types de phénomènes, malgré leurs liaisons, présentent quand même un certain nombre de différences sur lesquelles nous reviendrons ; ils peuvent d'ailleurs se produire indépendamment l'un de l'autre. D'autre part, d'un point de vue psycholinguistique, ils correspondent à deux processus différents. Ce point de vue apparaîtra plus clairement lorsque nous aurons détaillé les différentes caractéristiques de l'un ou l'autre. Aussi bien pour les corrections et erreurs que les disfluences, nous donnons dans un premier temps des définitions générales. Ensuite, nous présentons les principales études effectuées sur ce thème, et le cas échéant les modèles et définitions auxquels elles ont donné lieu.

#### 3.1. Erreurs et corrections

Au préalable, définissons le concept d'erreur. Cette définition doit pouvoir s'appliquer, non seulement à la production vocale, mais à tout champ d'activité. Pour cela, nous emprunterons la « définition de travail » que donne Reason (*op. cit.*, p. 9) :

**« l'erreur sera entendue comme un terme générique qui englobe toutes les circonstances dans lesquelles une séquence planifiée d'activités mentales ou physiques échoue à atteindre son but, et lorsque ces échecs ne peuvent être imputés à l'intervention d'un facteur aléatoire ».**

Notre démarche se veut double. Dans un premier temps, nous considérons les erreurs apparaissant dans le langage oral comme un cas particulier d'un ensemble auquel appartiennent également celles commises dans l'accomplissement de tâches telles que l'interaction homme-machine (Falzon (1989)), la conduite d'un véhicule, et plus exactement l'interaction avec un système automatique de copilottage (Bernsen *et al.* (2001)), le dialogue oral homme-machine (Bousquet (2004)), etc. Or, malgré la diversité

de leur contexte d'apparition, les erreurs présentent un nombre assez limité de patterns communs. Pour cette raison, nous jugeons nécessaire d'inscrire dans un cadre général l'étude de l'erreur, voire de l'échec communicationnel. Nous avons choisi de nous baser sur la description qui en est faite dans l'un des ouvrages de référence sur ce sujet : Reason (1990). Ce choix est motivé par le fait qu'il s'agit de l'un des rares travaux consacrés exclusivement aux différentes formes que peuvent prendre les erreurs. De plus, une perspective cognitive, essentiellement centrée sur celui qui fait l'erreur est adoptée, ce qui correspond à notre thème de recherche.

La deuxième étape de notre démarche consiste à s'intéresser aux erreurs de la production orale. Parmi les différents modèles existants, nous avons retenu le modèle d'architecture fonctionnelle de production orale de Levelt *et al.* (1999). Nous argumenterons le choix de ce modèle explicatif en section 3.1.2, où nous présenterons ce modèle.

### *3.1.1. Théorie générale de l'erreur*

J. Reason (1990) propose une catégorisation des erreurs selon divers critères. Nous retiendrons la principale taxonomie : les types d'erreurs sont déterminés en fonction du niveau d'exécution de la tâche où elles apparaissent. Reason distingue trois niveaux, qui sont relatifs à la temporalité d'accomplissement de la tâche. Ces trois niveaux sont les suivants :

- Planification des actions à accomplir (*planning*) : l'opérateur a une intention d'action par rapport à son environnement : parler, se déplacer, etc. Compte tenu des moyens dont il dispose et du contexte dans lequel il se situe, il va planifier la suite des actions nécessaires pour accomplir son intention. Par exemple, dans le cas de la production d'un énoncé, il s'agira de planifier ce qui va être dit. A ce niveau, une erreur consistera en une inadéquation entre les actions planifiées et leur résultat effectif sur l'environnement ;
- Mémorisation (*storage*) des actions planifiées, le temps de les mettre en œuvre : les actions sont exécutées séquentiellement (les unes à la suite des autres). Le niveau de mémorisation a pour rôle de conserver les actions qui ont été accomplies mais sont en attente d'exécution. Une erreur intervenant à ce niveau, si elle est remarquée par l'opérateur, entraîne la nécessité de recourir à nouveau à la phase de planification ; si elle n'est pas remarquée, elle se traduira par un arrêt de l'action entreprise ;
- Exécution (*execution*) des actions planifiées : comme son nom l'indique, ce niveau concerne l'exécution effective des actions planifiées et mémorisées, afin de réaliser l'intention de l'opérateur. Les erreurs à ce niveau sont provoquées par



une défaillance dans l'utilisation des organes effecteurs (par exemple, problème de coordination lors de la réalisation d'un geste).

Cependant, il y a d'autres critères entrant en compte pour une classification des erreurs. La réalisation d'une tâche peut notamment être analysée en fonction des compétences qu'elle nécessite. C'est ce qu'a entrepris Rasmussen (1983) qui propose une taxonomie des erreurs selon qu'elles sont relatives à :

- Des compétences (*skill-based*) : ce qui est appelé « compétence » correspond à l'utilisation par un humain de patterns d'actions et d'instructions pour l'exécution d'une tâche. Il s'agit de patterns qui sont tellement bien maîtrisés par l'opérateur qu'il n'a plus besoin d'y accorder une attention consciente. Un exemple classique est la conduite automobile ;
- Des règles (*rule-based*) : les règles dont il est question ici peuvent être formulées sous la forme « Si X, alors Y ». Elles sont utilisées comme heuristiques pour la résolution de problèmes, notamment dans le cadre de situations nécessitant une réflexion quant aux compétences à mettre en œuvre : le choix est guidé par les règles. Celles-ci sont déjà connues par l'opérateur ; leur application varie suivant les circonstances dans lesquelles il se trouve ;
- Des connaissances (*knowledge-based*) : contrairement aux deux catégories précédentes, celle-ci concerne les situations auxquelles n'a pas été confronté celui qui commet l'erreur. De ce fait, il ne peut compter ni sur ses compétences, ni sur des règles. Pour agir de manière appropriée, il doit avoir alors recours au raisonnement analytique ainsi qu'à l'utilisation de connaissances qu'il détient sur l'environnement dans lequel il se trouve (ou du système auquel il a affaire). Dans ce cas, l'erreur consiste en une mauvaise utilisation de ces connaissances, ainsi qu'en un recours à des données inappropriées.

Pour synthétiser, on dira que la première catégorie concerne uniquement des actions réalisées dans le cadre d'une activité régulière, tandis que les deux autres s'appliquent quant à elles à des tâches de résolution de problèmes (cf. Reason (*op. cit.*, p. 56 et sq.) pour d'autres traits distinctifs).

Une distinction importante est accomplie par Reason : celle d'erreur *latente*, opposée à l'erreur *immédiate*. Alors que l'erreur immédiate entraîne une conséquence visible sur le champ, l'erreur latente n'est pas remarquée immédiatement, et peut continuer à subsister avant de se déclarer en tant que telle, en raison de causes internes ou externes. Un exemple d'erreur latente dans le cadre du discours oral spontané est le quiproquo : un élément de la conversation semble, à tort, avoir été compris par les participants. Ces derniers ne se rendront pourtant compte de l'erreur que dans le cas où une contradiction apparaît

explicitement. On retrouve également ce même type d'erreur dans les situations de dialogue homme-machine où le sujet ne s'est pas aperçu d'une erreur du système de reconnaissance (cf. Bousquet *et al.* (2003)). Ces deux exemples illustrent la projection des catégories d'erreurs décrites par Reason à l'oral. A présent, nous allons nous pencher plus en détails sur les erreurs propres au discours oral spontané.

### 3.1.2. Erreurs et corrections dans le dialogue oral

Les travaux de Reason n'étant pas consacrés spécifiquement aux erreurs de performance du discours oral spontané, on n'y trouve pas une analyse fine de ce phénomène. Cependant, certaines observations sont intéressantes à relever.

Ainsi, l'auteur évoque la mise en évidence par des études expérimentales, au niveau cognitif, d'un module de contrôle (ou « éditeur ») de ce qui est dit, qui interviendrait aux différents niveaux du processus de production orale afin de comparer le résultat obtenu avec l'objectif poursuivi par le locuteur. Les études citées mettent en œuvre des expériences consistant à faire produire par les sujets des transpositions de lettres et/ou de mots<sup>9</sup>. Il a été alors observé que les erreurs relatives au niveau de traitement du lexique sont plus susceptibles d'échapper au module de contrôle que les autres. En effet, il a été constaté que les transpositions erronées faites par les sujets produisent significativement plus souvent des mots lexicaux que des *logatomes*<sup>10</sup>. Il en est déduit que

**“errors obeying lexical rules are more likely to evade the scrutiny of the “editor”, particularly if the latter is occupied elsewhere.” (Reason (*op. cit.*, p. 156)<sup>11</sup>.**

Ces observations suggèrent l'hypothèse de l'existence de plusieurs niveaux de traitement intervenant durant la production du discours oral, mais les travaux de Reason ne permettent pas de les identifier. De plus, Reason, ne s'intéresse qu'aux erreurs produites par des sujets « sains », alors qu'il est nécessaire selon nous, pour mieux comprendre la production langagière, de s'intéresser à l'explication des troubles langagiers (cf. Nespoulous *et al.* (2004a)). Ces constats nous invitent à adopter une approche psycholinguistique expérimentale, complémentaire à celle de Reason qui est essentiellement cognitive. Il s'agit du modèle de Levelt (Levelt (1999)).

Pour étudier la production de la performance de messages oraux, les psycholinguistes ont recours à des modèles visant à représenter les différents niveaux qui interviennent dans ce processus. L'ensemble de ces niveaux constitue une architecture fonctionnelle (par

---

<sup>9</sup> Les sujets étaient « enclins » à ces erreurs en faisant précéder les mots cibles par des éléments susceptibles d'augmenter la possibilité de telles erreurs.

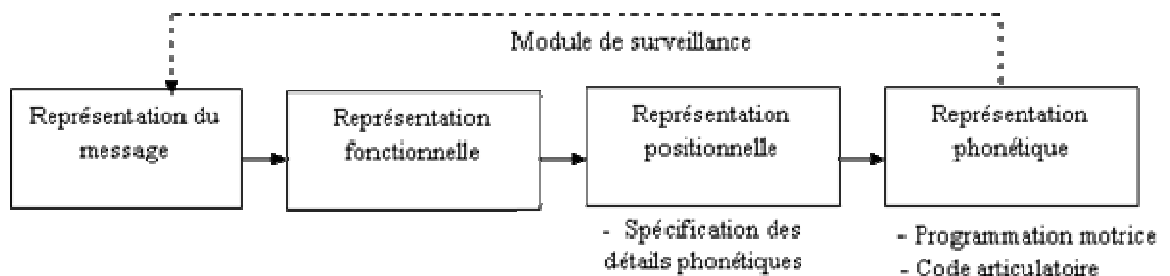
<sup>10</sup> Un *logatome* est un mot « artificiel » n'appartenant pas au lexique d'une langue donnée, mais construit selon les règles phonologique de celle-ci. Il peut donc donner l'impression d'appartenir à cette langue. Les logatomes sont souvent utilisés dans les expériences neuro- ou psycho-linguistiques.

<sup>11</sup> Notre traduction : « les erreurs obéissant à des règles syntaxiques sont plus susceptibles d'échapper à la surveillance du module de contrôle, particulièrement si ce dernier est occupé par ailleurs »

opposition à neurologique : il s'agit de rendre compte de l'aspect psycholinguistique et non de celui organique). L'utilisation de ce type de modèle pour l'étude des troubles de la parole consiste généralement à déterminer à quel niveau du modèle intervient le problème. Ce diagnostic est effectué au moyen de tests de production qui contrôlent un niveau en particulier. Suite à cela, des hypothèses sont formulées quant à la nature et l'étendue du trouble. Elles peuvent éventuellement être confirmées par l'utilisation d'analyses neurologiques et d'imagerie cérébrale, etc.

Deux types d'approches président à la constitution d'une architecture fonctionnelle. On trouve d'une part l'approche dite *connexionniste*, qui décrit les différents niveaux de traitement de l'énoncé à produire. Ceux-ci interagissent entre eux de manière non-hiérarchisée. A l'opposé, on retrouve les approches *linéaires* ou *sérielles*, dans le sens où les niveaux de traitement sont ordonnés en fonction de leur position lors de la production du message. Les deux principaux représentants en sont les modèles de Garrett (1980) et de Levelt (1999). Ce dernier est inspiré du modèle de Garrett qu'il a contribué à développer.

Le modèle de Levelt a pour but de schématiser l'ensemble du processus de production orale, depuis la représentation du message jusqu'à son articulation par les organes phonatoires. Il consiste en la succession de différentes phases, chacune correspondant à un niveau spécifique de traitement.



**Figure 2 :** Présentation simplifiée des principaux niveaux du modèle de Garrett

La Figure 2 ci-dessus en présente les principaux niveaux. Le premier, *représentation du message*, correspond à la recherche des éléments lexicaux les plus appropriés pour rendre compte du sens du contenu linguistique du message. Le second consiste à produire une représentation phonologique correcte des lexèmes en question, et leur assigner une conjugaison et une place dans l'énoncé qui respectent les règles syntaxiques. Il reste ensuite à spécifier les détails phonétiques (troisième niveau : *représentation positionnelle*) et enfin à effectuer la planification du code articulaire (*représentation phonétique*). Signalons enfin l'intervention d'un module de surveillance qui vérifie la conformité de la

chaîne phonétique produite avec le sens que le locuteur souhaitait exprimer. Son rôle correspond au contrôle décrit dans Reason.

Nous adoptons le cadre fourni par ce modèle. En effet, il est spécialement approprié à l'établissement de points de comparaison entre les études sur les erreurs langagières pathologiques et celles commises par des sujets « sains ». Cela est dû à son origine, puisqu'il a été établi à partir de l'observation des erreurs de performances de deux catégories de population, d'abord « normale » mais aussi aphasique<sup>12</sup> dans un deuxième temps (Nespoulous (2004a, p. 9 et *sq.*)). De nombreux auteurs ont utilisé ce modèle plus pertinent pour rendre compte de la production du discours et de ses aléas. Ainsi, Rossi démontre que les relations entre les différentes fonctions langagières doivent nécessairement être asymétriques, et conclut que

**« un modèle connexionniste n'était pas apte à construire ou manipuler les symboles du niveau cognitif » (Rossi et al., (1998, p.88)).**

Nous verrons plus loin dans ce chapitre d'autres utilisations de ce modèle dans des études sur les disfluences.

### **3.2. Les disfluences**

#### *3.2.1. Premières définitions*

Comme l'indique leur étymologie, les disfluences correspondent à toute interruption ou perturbation de la *fluence*, c'est-à-dire du cours « normal » de la production orale spontanée. Contrairement à d'autres phénomènes spécifiques au langage oral spontané (comme les clitics par exemple), leurs occurrences ne sont pas produites volontairement par le locuteur<sup>13</sup>. Le terme générique de disfluences recouvre en fait un ensemble disparate de phénomènes et présente de multiples modes de manifestations que nous présenterons plus bas. Contrairement aux erreurs, elles n'apparaissent qu'à l'oral spontané, et non à l'écrit<sup>14</sup>.

---

<sup>12</sup> Rappelons que l'aphasie est une atteinte d'une des aires cérébrales gérant le langage chez l'homme. De ce fait, les symptômes qu'elle provoque sont souvent étudiés en psycholinguistique, notamment pour mieux comprendre les erreurs de performances..

<sup>13</sup> Comme nous le verrons plus bas, le caractère « involontaire » des disfluences est discuté par certains auteurs.

<sup>14</sup> Ce dernier point peut lui aussi être contesté, ou à tout le moins discuté. En effet, la principale raison pour laquelle on n'observe pas de disfluences à l'écrit est le fait que le lecteur n'a pas accès au texte en cours de production, mais uniquement au résultat final. L'auteur (assimilable au locuteur de l'oral spontané) a donc tout loisir de corriger et réviser son texte pour en obtenir une version conforme aux standards. Cette possibilité n'est pas disponible à l'oral, comme nous l'avons vu en début de chapitre.

Deux observations peuvent cependant être faites. D'une part, il reste souvent des scories du processus de rédaction de l'écrit, même après relecture de celui-ci. On parle parfois à ce propos de « coquille » : fautes d'orthographe, mot répété, etc. D'autre part, dans certains contextes, il est nécessaire d'avoir accès au texte écrit en cours de production. Nous pensons particulièrement au développement de systèmes d'assistance à la saisie de texte, destinés notamment aux personnes souffrant d'handicap moteur. Ces systèmes ont deux objectifs. D'une part compléter automatiquement les mots avant la fin de leur production, afin d'éviter la fatigue d'une saisie trop longue. D'autre part, corriger les fautes de frappes. Pour atteindre ces deux buts, il

Ce phénomène est lui aussi abondamment traité dans la littérature. Le fait qu'il recouvre différents types de manifestations pose problème : en effet, la typologie et la terminologie employées varient d'un auteur à l'autre, comme le remarquent souvent les auteurs eux-mêmes (cf. par exemple Shriberg (1992, p.10)). Il est alors difficile de se situer par rapport aux unes et autres définitions, ou même d'en proposer une catégorisation « générique ».

Malgré tout, nous allons tenter de dresser une liste sommaire et générale des principales catégories de phénomènes, à partir des différentes études existantes. L'objectif de cette liste est de servir de point de départ et de « fil rouge » pour une présentation plus approfondie. Elle se compose des 5 éléments suivants :

- Les hésitations : le classique « euh » ; plus généralement tout son ne correspondant pas à un mot de la langue ;
- Les silences : cette notion est assez subjective. En effet, elle nécessite de fixer une borne à partir de laquelle estimer que la période de silence entre deux mots n'est plus « normale » mais disfluente. Cela peut être fortement variable selon le contexte, le locuteur, etc. ;
- Les répétitions : toute répétition d'un ou plusieurs mots. Comme nous le verrons, cette catégorie est bien plus difficile à déterminer qu'il n'y paraît au premier abord ;
- Les auto-corrections : toute correction faite par le locuteur lui-même, consistant en une interruption de la production orale en cours. Cette catégorie n'inclut évidemment pas les corrections effectuées sous une forme discursive. Par exemple, « Le train pour Paris non Perpignan » comporte une auto-correction au sens disfluent du terme, contrairement à « Le train pour Paris. Excusez-moi je veux dire Perpignan ». Il s'agira par contre d'une erreur et d'une correction au sens que nous avons exposé en 3.1.2 ;
- Autres : cette catégorie regroupe tous les autres phénomènes disfluents qui ne rentrent pas dans une catégorie accessible immédiatement au « sens commun ». Au fur et à mesure de notre revue de l'état de l'art et de nos propres analyses, cette catégorie sera raffinée et précisée.

La suite de ce chapitre va être consacrée à une description bien plus fine de chacune de ces catégories. Pour cela, nous allons présenter les principales études qui ont été menées

---

est nécessaire de modéliser les fautes en cours de production (voir par exemple Boissière *et al.* (2007), et Bouraoui *et al.* (2007)). Cette tâche se rapproche de la démarche que nous entreprenons pour l'oral. Mais elle s'en démarque malgré tout par un certain nombre de différences, notamment la modalité de production (oral VS écrit au clavier). C'est en raison de ces différences que cette discussion se situe en note de bas de page et non dans le corps principal du chapitre.

sur le sujet. Evidemment, la définition de « principale » est ici subjective. Nous avons effectué le choix de ces études selon deux critères non exclusifs. Il s'agit d'une part de l'impact qu'elles ont eu sur des travaux ultérieurs sur des disfluences (par exemple, mise en place d'un cadre méthodologique réemployé plus tard). D'autre part, le deuxième critère de choix est la proximité avec les objectifs de cette thèse, à savoir la détection automatique des disfluences dans le langage oral spontané. C'est lors de la présentation de ces différents travaux et de leurs résultats que nous détaillerons les caractéristiques des différentes catégories de disfluences. Le cas échéant, nous donnerons les distinctions existant selon les auteurs, et leurs motivations).

### 3.2.2. Blanche-Benveniste

Les travaux de C. Blanche-Benveniste sont souvent utilisés dans les études ultérieures sur les disfluences. En effet, ils fournissent un cadre pertinent de représentation et d'interprétation de ces phénomènes.

Blanche-Benveniste n'effectue pas une classification précise des différentes catégories de disfluences, étant donné que ce n'est pas l'objectif principal du travail effectué. Le chapitre 2 de Blanche-Benveniste (1997, p. 45-49) présente néanmoins deux catégories de phénomènes assimilables à des disfluences : les *inachèvements* et les *auto-corrections*. La première catégorie s'applique uniquement aux phrases inachevées ; ne sont donc pas concernées des unités plus petites de l'énonciation, telles que les mots. Les auto-corrections, quant à elles, ne sont pas définies formellement, mais les exemples donnés montrent que les phénomènes considérés correspondent à la première définition que nous donnons ci-dessus.

Selon Blanche-Benveniste, ces phénomènes peuvent être expliqués dans la perspective des axes syntagmatiques et paradigmatiques que nous avons présentée en 2.1. En effet, des phénomènes tels que les répétitions ou les auto-corrections correspondent à des *entassements* sur l'axe paradigmatiques. Ces entassements sont remarqués (plus précisément : audibles) à l'oral en raison de la propriété de linéarité de celui-ci. Ainsi, l'énoncé « alors *les* parents, euh *certain*s parents, viennent »<sup>15</sup> comportant une auto-correction, sera représenté ainsi :

Alors les parents euh  
certaines parents viennent

On voit que la compréhension et la description des disfluences est facilitée. De plus, un traitement systémique des disfluences est ainsi permis, au même titre que l'ensemble des autres phénomènes propres à l'oral.

---

<sup>15</sup> Exemple cité dans Blanche-Benveniste (*op. cit.*, p. 49, souligné dans le texte original).

A partir de cette modélisation Blanche-Benveniste propose une nouvelle théorie pour rendre compte de ces phénomènes, qu'elle appelle la *macro-syntaxe*.

### 3.2.3. Shriberg

La thèse de E. Shriberg (1994) est un des travaux fondateurs dans les théories modernes sur les disfluences ; notamment grâce à l'approche d'analyse systématique qui est entreprise, ainsi que par le modèle qui est employé, et repris (même si c'est parfois pour être critiqué) par de nombreuses études ultérieures sur ce thème.

Voyons à présent plus en détail en quoi consiste ce modèle (Shriberg (1994, p. 7-9, et chapitre 4, spécialement p. 62-75)). Son but est de rendre compte de manière précise de l'occurrence d'une disfluente, et de la manière dont elle affecte son environnement dans l'énoncé. Il est fondé sur le modèle de Levelt, que nous avons présenté plus haut (section 3.1.2). Nous reproduisons ci-dessous la figure utilisée par l'auteur pour présenter son modèle. Le cas présenté ici est celui d'une disfluente relative à la correction par le locuteur de sa production. Mais comme nous le verrons plus bas, il s'applique également à d'autres types de disfluences.

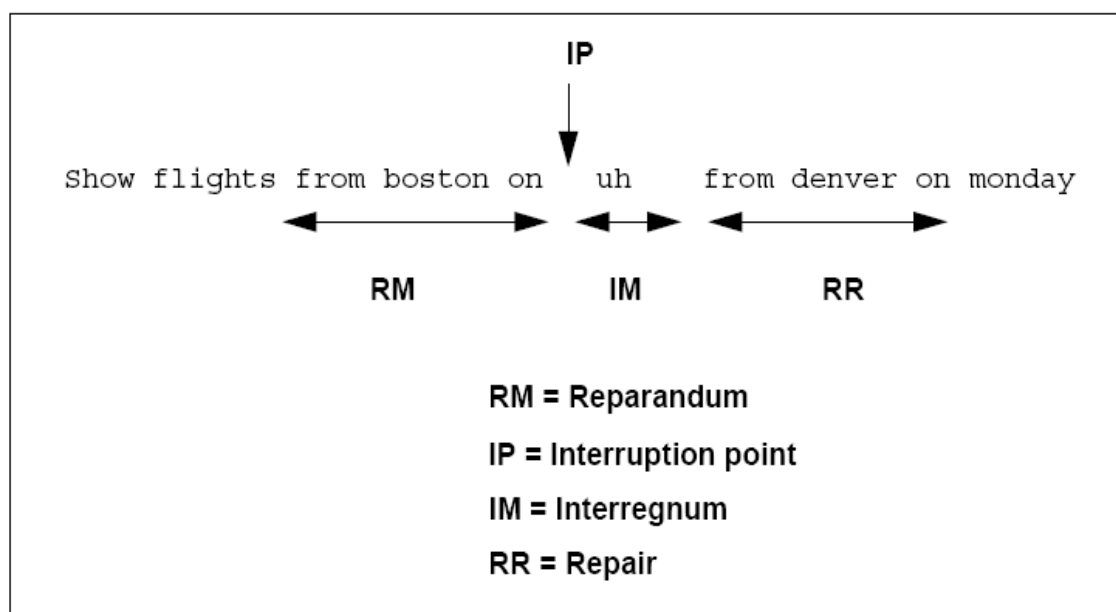


Figure 3 : Modèle des disfluences de Shriberg

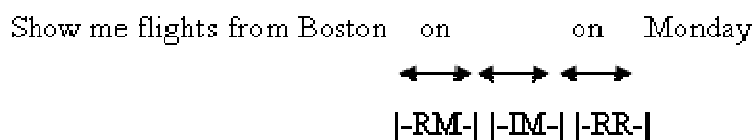
La disfluente est ici le « uh », correspondant à un « euh » en français. Avant son occurrence se situe le *Reparandum* (RM). Il s'agit de la zone de l'énoncé dont le contenu va être corrigé ou reformulé.

Après le *Reparandum* se trouve la zone appelée *Interregnum* (IM) : celle-ci correspond à l'intervalle situé entre le *Reparandum* et la suite de l'énoncé qui constituera en la

correction ou la reformulation. C'est dans l'*Interregnum* que peut apparaître la disfluence. Mais Shriberg précise que cette zone est de toute façon présente même en l'absence de disfluences : elle correspond à l'intervalle de temps durant lequel le locuteur procède à la re planification de son énoncé. Enfin, l'*Interregnum* ne correspond pas systématiquement à une pause de longue durée, et peut être aussi bref qu'une pause « normale » entre deux mots.

Enfin, à la suite de l'*Interregnum* se trouve la phase *Repair* (RR). Elle comporte la correction ou reformulation de la partie de l'énoncé située dans le *Reparandum*.

Comme nous le disions plus haut, la structure et les différentes zones du modèle que nous venons de décrire s'applique non seulement aux corrections, mais également aux autres types de disfluences. La différence résidera dans le fait que certaines des zones seront « remplies » ou non par des éléments de l'énoncé. Par exemple, la répétition d'un mot sera représentée ainsi (exemple inspiré par Shriberg (1994, p. 9)) :



**Figure 4 :** Exemple d'application du modèle de Shriberg

En termes de classification, l'auteur se fonde, en apportant des modifications, sur un système d'étiquetage des disfluences conçu au *Stanford Research Institute (SRI)*. Les catégories de disfluences prises en compte dans ce travail sont essentiellement décrites et détaillées dans Shriberg (*op.cit.* p. 57- 69).

Etant donné que le travail exposé consiste à étiqueter les disfluences, la description qui en est faite regroupe aussi bien les disfluences elles-mêmes, que leur mode de manifestation et les délimiteurs utilisés pour l'étiquetage. Nous ne présentons ici que les seules catégories de disfluence.

#### 3.2.3.1. Les répétitions ("repeated words")

Désigne tout mot qui est répété de manière identique avant et après le point d'interruption.

#### 3.2.3.2. Les pauses remplies ("filled pauses")

Il n'y a pas à proprement parler de définition pour cette catégorie. Les éléments concernés sont « uh » et « um ».



3.2.3.3. *Les termes explicites d'édition ("explicit editing terms")*

Là encore pas de définition, mais les exemples donnés sont clairs : « no », « oops », « i'm sorry », etc. Ces exemples montrent que les termes d'édition ont pour rôle de signaler explicitement à l'interlocuteur que le locuteur s'apprête à corriger ce qu'il vient de dire.

3.2.3.4. *Les marqueurs de discours ("discourse markers")*

Comme pour la catégorie précédente, pas de définition, mais des exemples de ces marqueurs : « well », « you know », « like ». Contrairement à l'utilisation des termes d'édition, le locuteur signale qu'il est en train de chercher à poursuivre sa production.

3.2.3.5. *Les fragments de mot ("word fragments")*

Une définition informelle en est faite p. 26

**“speech may be cut off mid-word, even mid syllable, yielding what will be referred to in this thesis as a “word fragment” or simply a “fragment” at the right edge of the RM”**<sup>16</sup>

3.2.3.6. *Insertions ("inserted word"), suppressions ("deleted word"), substitutions ("Word in substituted string" / "Substituted-string fragment")*

Dans la thèse de Shriberg, ces trois phénomènes sont présentés séparément les uns des autres. Nous les regroupons dans la même rubrique par commodité de présentation, et parce qu'ils relèvent tous de ce que nous avons appelé plus haut l'auto-correction (bien que ce terme ne soit pas employé). On peut en juger par les exemples donnés pour chaque phénomène, et que nous reproduisons ci-dessous.

- Insertion : désigne l'apparition d'un mot dans la zone de *Repair*. Par exemple :  
please give me fares **round trip** fares from pittsburgh
- Suppression : il y a suppression lorsqu'un mot qui était présent dans le *Reparandum* est absent dans la zone de *Repair*. L'exemple donné est le suivant :  
flights from boston **to** on friday  
L'auteur signale que ce type de phénomène est souvent dénommé *false start* ou *fresh start* (autrement dit, « faux départ » en français) dans d'autres systèmes de classification.
- Substitution : il y a substitution lorsqu'un ou plusieurs mots du *Reparandum* sont remplacés par un ou plusieurs mots dans le *Repair*. Pour qu'il y ait

---

<sup>16</sup> Notre traduction : « le discours peut être interrompu à mi-mot, même à mi-syllabe, provoquant ce à quoi nous référerons dans cette thèse comme un "fragment de mot" ou simplement un "fragment", à droite du RM ».

substitution, il faut également qu'il y ait correspondance sémantique et syntaxique entre le ou les mots substitués et la substitution. Par exemple :

does United flight 201 serve a *sn(ack)-* breakfast

Dans cet exemple, "snack" est remplacé par "breakfast"; il y a bien correspondance syntaxique et sémantique entre les deux mots.

### 3.2.4. Candea

Le principal travail de M. Candea auquel nous nous référerons ici est sa thèse de doctorat : Candea (2000). Les disfluences sont considérées à la fois sous l'angle de la psycholinguistique et de la linguistique.

Le corpus sur lequel le travail est effectué comprend 13 histoires racontées oralement par des enfants. Sa durée est de 70 minutes et 25 secondes.

Cinq principales catégories de disfluences sont étudiées dans la thèse. Pour chacune, nous donnons les définitions faites par l'auteur, et ajoutons quelques commentaires.

#### 3.2.4.1. Les pauses silencieuses

Elles sont définies comme :

**« une interruption significative de toute émission sonore à l'intérieur d'une prise de parole. » (Candea, *op.cit.*, p. 21).**

L'auteur souligne les différentes nuances à apporter à cette définition. D'abord le fait que des pauses interviennent « normalement » dans le cadre de toute production orale : que ce soit les reprises de souffle, le moment intermédiaire entre la fin d'une prise de parole d'un dialoguant et celle du suivant, etc. Comment alors distinguer une pause silencieuse disfluente d'une « normale » ? Le critère principal est la durée : une pause sera considérée comme disfluente si elle dure 20 centièmes de seconde (cs) ou plus. Ce choix est motivé par diverses études existantes, dont nous renvoyons la référence à Candea (*op.cit.* p. 22). Il n'y a par contre pas de durée maximale au-delà de laquelle une pause n'est plus considérée comme disfluente.

#### 3.2.4.2. Les euh dits d'hésitation

Candea les définit ainsi :

**« Voyelle prononcée [ə, oe, ø voire ε̃ ou oem] pouvant être soit insérée en épenthèse en finale de mot et formant une syllabe supplémentaire, soit prononcée indépendamment, avant ou après un mot. Dans ce dernier cas elle peut être séparée du mot précédent et éventuellement du mot suivant par une interruption du signal sonore plus ou moins longue, ou bien, dotée d'une montée de l'intensité qui en fait une syllabe autonome par**

**rapport à la séquence précédente. Cette voyelle représente l'indice le plus largement connu pour marquer conventionnellement la recherche de formulation en français oral. » (Candea, *op.cit.*, p. 24)**

Comme elle constitue une entité en tant que telle, il n'y a pas de seuil minimum nécessaire pour caractériser l'hésitation.

#### 3.2.4.3. *Les allongements vocaliques*

En raison de leurs propriétés phonétiques, seules les voyelles peuvent faire l'objet d'allongements, qui sont définis comme suit :

**« Tout allongement vocalique anormal en position finale de mot ou d'amorce de mot, présentant un contour plat et bas ou très légèrement descendant, représente un allongement marquant le travail de formulation en cours. Une voyelle commence à être anormalement allongée lorsque sa durée se situe entre 18 et 22 cs, selon les locuteurs. » (Candea, *op.cit.*, p. 26)**

#### 3.2.4.4. *Les répétitions*

Comme nous le signalions en 3.2.1, il s'agit d'une catégorie particulièrement difficile à définir. Commençons par la première définition donnée par l'auteur :

**« Nous appellerons 'répétition' destinée à marquer le travail de formulation toute répétition à l'identique, en contiguïté dans la chaîne sonore d'un son, d'une syllabe, d'un mot ou d'une amorce de mot, de plusieurs syllabes ou de plusieurs mots, sans aucune valeur sémantique. » (Candea, *op.cit.*, p. 28)**

De par sa définition, l'auteur exclut deux types de répétitions. D'une part celles dans lesquelles la partie répétée diffère, même légèrement, de sa première occurrence. La justification est que ce type de phénomène est catégorisé comme appartenant à une sous-catégorie des autocorrections ; nous en reparlerons donc plus bas. D'autre part, l'usage volontaire de répétitions non disfluentes, par exemple dans le cas d'exagération stylistique, ou pour raisons syntaxiques (par exemple « nous nous sommes »). La « valeur sémantique » évoquée dans la définition renvoie à ce dernier cas.

La répétition disfluente est constituée de deux parties. D'une part le *répétable*, c'est-à-dire l'unité faisant l'objet de la répétition. D'autre part, le bloc du ou des *répété* : il comprend de 1 à  $n$  répétitions du répétable ; chaque occurrence est appelée répété.

#### 3.2.4.5. Les autocorrections

L'auteur en distingue deux types, les autocorrections « immédiates » et les autocorrections « plus complexes ou faux départs ». La définition du premier type est la suivante :

**« dans une séquence parlée de type X Y, nous considérons que la partie Y est une autocorrection immédiate de X**

**- si Y est prononcé pour remplacer X, et**

**- si Y corrige un seul et unique trait phonétique ou morphologique de X.**

**Nous incluons dans les autocorrections immédiates, en raison de leur durée très brève, les cas où Y opère un changement de classe par rapport à X, à condition que X et Y soient des mots outils et qu'ils occupent la même place syntaxique par rapport à ce qui précède X (comme dans l'exemple est-il vrai : que tu que : que ta fille est capable ..., ligne 294, où le pronom tu est remplacé par le SN ta fille qui occupe la même place, celle du sujet, dans la subordonnée introduite par que). » (Candea, *op.cit.*, p. 29)**

Le dernier paragraphe de cette définition est relatif au cas, que nous évoquons plus haut, de répétitions qui ne sont pas complètement identiques. L'auteur précise également le point fondamental qui fait la différence entre une répétition et une autocorrection immédiate :

**« L'autocorrection immédiate s'apparente à la répétition dans la mesure où elle représente une reprise de tout un faisceau de traits syntaxiques, morphologiques et phonétiques... sauf un. » (Candea, *op.cit.*, p. 30 ; souligné dans le texte original)**

Par contre, tous les types de faux-départs ne sont pas pris en compte par l'auteur. Ils sont en effet assimilés à des techniques dites de « retouches stylistiques ». Elles sont utilisées notamment à des fins d'insistance sur un point particulier d'un énoncé.

#### 3.2.4.6. Autres observations sur les disfluences

Deux points sont notables dans le travail de Candea. Le premier est la constatation que certaines disfluences n'apparaissent pas « au hasard » mais dans certaines configurations lexico-syntaxiques données. Une formulation synthétique – et modérée – de cette constatation est la suivante :

**« En résumé, il nous semble que s'il a indéniablement été prouvé que les autocorrections et faux départs sont soumis à certaines contraintes syntaxiques, il n'a pas encore été prouvé que ces phénomènes pourraient être décrits par un modèle syntaxique simple et unique. » (Candea, *op.cit.*, p.350)**

Cette constatation est corroborée par d'autres travaux, notamment Henry (2002 p. 474-475) et Blanche-Benveniste (2003).

Le deuxième point est le fait que certaines disfluences, notamment les allongements et les hésitations, doivent être analysées et interprétées de manière plus fine que des « simples » perturbations de l'oral. Candéa propose plutôt de les considérer comme des « marques du travail de formulation ». De ce fait, elles acquièrent une valeur pour l'analyse de la planification de la production orale.

Cette interprétation est rejointe par celles d'autres auteurs, parmi lesquels Clark et Fox Tree (voir par exemple Clark *et al.* (2002)). Ils proposent en effet de considérer que les disfluences sont des indicateurs utilisés par le locuteur pour indiquer ses intentions (par exemple, qu'il cherche une formulation particulière, qu'il n'a pas fini de parler, etc.). Leur analyse les conduit même à estimer que les hésitations sont des mots (du moins en anglais, puisqu'il s'agit de la langue sur laquelle ont porté leur étude), appartenant à la catégorie des interjections.

#### 3.2.5. Henry et Pallaud

Nous nous référerons ici à différents articles écrits par l'un et/ou l'autre de ces auteurs, relatifs à un ou plusieurs des phénomènes de disfluence. Il s'agit des articles référencés dans notre bibliographie : Henry (2002), Henry *et al.* (2003), Henry *et al.* (2004), Pallaud *et al.* (2004). L'immense majorité des corpus utilisés sont des « sous ensembles » de *Corpaix*, constitué par le *Groupe Aixois de Recherche en Syntaxe (GARS)*, et/ou du *Corpus de référence de français parlé*, collecté par l'équipe *DELIC (DEscription Linguistique Informatisée sur Corpus ; ex-GARS)*. Ces corpus sont constitués de dialogues oraux spontanés enregistrés dans différentes situations, que ce soit des entretiens ou conversations « libres » ou des tâches un peu plus orientées, comme par exemple des demandes de chemin, des descriptions, etc. Pour plus de détails sur le contenu de ces corpus et leurs points communs et différences, nous renvoyons à l'ensemble des références citées ci-dessus, et plus particulièrement à Pallaud *et al.* (2004, p. 849-850). Il nous semble par contre pertinent, dans le cadre de notre perspective de travail, de mentionner d'abord le fait que ces corpus ont pour objectif de représenter « un français parlé qu'il serait possible de qualifier d'un « usage général et courant » (Pallaud *et al.*, (*op. cit.*, p. 849)). Ce point est en effet important puisque que la nature de la tâche accomplie par les locuteurs d'un corpus influe grandement sur ces caractéristiques linguistiques, comme nous le montrerons dans le chapitre 3. D'autre part, quelques considérations statistiques s'imposent pour donner un point de comparaison inter-études. Le *Corpus de référence de français parlé* est constitué de 1 000 382 mots, produits par 794 locuteurs différents, dans 283 situations différentes.

Dans leurs analyses de corpus, les auteurs distinguent les catégories suivantes de disfluences.

### 3.2.5.1. Les hésitations

Ce terme désigne les pauses, « silencieuses » ou « remplies ». Une pause sera dite silencieuse lorsqu'aucun son n'est émis par le locuteur pendant un certain temps. Par opposition, une pause remplie est une production vocale qui ne correspond pas à ce que voulait le locuteur. Cela peut être par exemple le classique « euh », ou tout autre son compris entre la fin d'un mot donné et le début du suivant. L'une des questions primordiales compte tenu de cette définition, au moins pour la pause silencieuse, est évidemment la durée à partir de laquelle on a affaire à une hésitation et non à une pause normale entre deux mots. De ce point de vue, les auteurs choisissent, comme Candéa et d'autres, une durée de 200 millisecondes (autrement dit, 20 cs). Comme on le voit, le terme « hésitation » recouvre trois catégories qui sont distinguées chez Candéa, à savoir le « euh » d'hésitation, les pauses silencieuses et les allongements vocaliques.

### 3.2.5.2. Les répétitions

C'est le phénomène qui est le plus étudié par ces deux auteurs. Il est d'abord établi, comme chez d'autres auteurs, une distinction, entre les répétitions d'origine volontaire (quelle qu'en soit la raison : effet de style, volonté d'exprimer une exagération, etc.) et les répétitions dites « de performance » car elles sont provoquées par un problème à ce niveau de la production orale. Les répétitions de performance peuvent affecter des unités de l'énoncé de granularités différentes : fragments de mots (cf. plus bas), mots, syntagmes. Les auteurs reprennent la classification faite par Candéa (cf. plus haut, section 3.2.4.4) entre « répétable » et « répété ».

Les fragments de mot (aussi appelés « amorces » dans Henry *et al.* (2004)) sont définis comme des mots dont la production est stoppée précocement, avant leur fin « normale ». Trois catégories différentes sont distinguées, selon la manière dont cette disfluence se manifeste :

- Fragments de mot inachevés : malgré la présence du mot incomplet, la phrase continue comme si le mot avait été prononcé jusqu'au bout ;
- Fragments de mot complétés : le mot initialement laissé inachevé est redit, cette fois-ci jusqu'à sa fin ;
- Fragments de mot modifiés : contrairement à ce qui se produit pour la catégorie précédente, le mot inachevé n'est pas repris, mais est remplacé par un autre.

Les auteurs remarquent aussi que, d'une manière générale, les différentes disfluences apparaissent rarement seules, mais en combinaison les unes avec les autres. Par exemple il peut y avoir des répétitions de fragments de mots ou d'hésitations, des conjonctions de

plusieurs catégories, etc. C'est même un cas fréquent, puisque par exemple 77% des répétitions sont associées avec au moins une pause (Henry *et al.* (2004, p. 262)).

#### 3.2.6. Kurdi

L'ensemble de la thèse de Z. Kurdi (Kurdi (2003)) est consacré à la modélisation des différentes difficultés et problèmes pouvant advenir dans le cadre du dialogue oral spontané, dans une perspective de Traitement Automatique.

Elle est basée sur un corpus de négociations (en anglais) de transport de marchandises par train comprenant 52 000 mots. Ce corpus est donc relativement orienté par la tâche à accomplir.

Z. Kurdi, contrairement à la majorité des autres études de la littérature sur le sujet, n'utilise pas le terme de disfluences. Il trouve en effet que ce terme réfère trop à des caractéristiques prosodiques ou phonétiques, alors que le phénomène désigné ainsi devrait être décrit de manière plus générale et précise (Kurdi (*op. cit.*, p. 20)). Il préfère donc employer le terme de « extragrammaticalité », lui-même proposé dans Carbonell *et al.* (1983).

A partir de là, la classification proposée par l'auteur comporte deux principales catégories : les *extragrammaticalités lexicales* et les *extragrammaticalités supralexicales*. La différence entre les deux catégories est le fait que la première concerne uniquement, comme son nom l'indique, des unités discrètes assimilables à des lexèmes, alors que la seconde est relative à un niveau supérieur aux dites unités<sup>17</sup>. Chacune de ces catégories recouvre elle-même un certain nombre de phénomènes différents, que nous allons à présent présenter brièvement.

##### 3.2.6.1. Les extragrammaticalités lexicales

Les phénomènes appartenant à cette catégorie sont dénommés *pauses*, *mots oraux*, *mots incomplets*, et *amalgames*. Comme dans l'école d'Aix, les *pauses* peuvent être « simples » (aucun signal n'est émis) ou « remplies » (même définition que celle que nous avons présentée plus haut) Dans le cas des pauses remplies, l'auteur emploie également le terme d'hésitation.

Les *mots incomplets*, comme leur nom l'indique, correspondent à des mots dont le locuteur arrête la production avant leur fin. Comme le fait remarquer Kurdi, cette sous-catégorie est en général ignorée par les systèmes de reconnaissance de la parole, et n'est donc pas

---

<sup>17</sup> Cette classification est sans doute faite par analogie avec les travaux linguistiques d'André Martinet, présentés notamment dans Martinet (1960), et faisant une distinction entre deux « articulations » du langage : le niveau segmental et suprasegmental. Cependant, les phénomènes désignés par ces deux termes respectifs ne sont pas exactement les mêmes que chez Kurdi ; par exemple, le niveau suprasegmental correspond, chez Martinet à tout ce qui concerne la prosodie, l'accent, etc.

utilisable lorsque ces systèmes sont employés. Les *mots oraux* désignent tout mot dont l'usage est spécifique à l'oral (par exemple « ouais » en français, l'équivalent anglais étant « yeah », etc.) ; de plus leur emploi, même à l'oral, est caractéristique d'un certain registre de langue qui peut parfois faire l'objet de correction (cf. par exemple « ouais » au lieu de « oui », cité par Kurdi (*op. cit.*, p. 24)). Enfin, les *amalgames* correspondent aux différentes contractions utilisées à l'oral ; par exemple « I'll » pour « I will » en anglais, ou « chuis » pour « je suis » en français<sup>18</sup>.

### 3.2.6.2. Les extragrammaticalités supralexicales

Les termes utilisés pour phénomènes appartenant à cette deuxième catégorie (répétitions, autocorrections, faux-départs, incomplétude) sont identiques à ceux utilisés dans la majorité de la littérature, à l'exception de « incomplétude ». Par contre, leur définition est parfois différente, comme nous allons le voir en détaillant chacun d'entre eux.

- Les répétitions : la définition donnée est proche de celles des autres auteurs que nous avons cités auparavant. Il y a notamment prise en compte de la différence entre une répétition employée volontairement à des fins stylistiques ou autres, et une répétition disfluente. Cependant, l'auteur fait une distinction qu'il est à notre connaissance le seul à utiliser, puisqu'il considère que

**« La répétition est définie sur des critères purement morphologiques. Par conséquent, la formulation et la paraphrase d'un énoncé ou d'un segment (où l'on répète deux segments qui ont le même sens) ne sont pas considérées comme étant des répétitions : (...) ce serait un vol Paris Delhi plus un vol un vol intérieur. » (Kurdi, *op. cit.*, p. 24, souligné dans le texte original)**

Dans cet exemple, le syntagme « un vol intérieur » est considéré comme étant une paraphrase du syntagme précédent « un vol ». Conformément à la définition donnée, ce n'est pas une répétition ;

- Les autocorrections, contrairement aux répétitions, consistent :

**« à remplacer un mot ou une série de mots par d'autres afin de modifier ou corriger le sens de l'énoncé. L'autocorrection n'est pas complètement aléatoire et porte souvent sur un segment qui peut compter un ou plusieurs syntagmes (Core, 1999), c'est pourquoi elle**

---

<sup>18</sup> Ces contractions se font généralement selon le principe du moindre effort (certains parlent « d'économie linguistique »), mis en évidence par Martinet (cf. par exemple Martinet (1960)), selon lequel tout locuteur, consciemment ou non, tendra à réduire autant que possible la durée de sa production, en raison de la fatigue occasionnée.

Elles peuvent également être provoquées par des phénomènes d'ordre phonétique ou phonologique tels que les *assimilations* par exemple (comme c'est le cas pour « je suis » prononcé « chuis » : assimilation régressive du « s » sur le « j »).



**est fréquemment accompagnée par une répétition partielle du segment corrigé. » (Kurdi, *op. cit.*, p. 24)**

Un exemple d'auto-correction correspondant à cette définition est le suivant (dans lequel la disfluenza est soulignée) :

Oui : j'ai la j'ai les pages Web oui

On observe que le syntagme faisant l'objet de la correction (« j'ai la ») est presque identique à celle-ci (« j'ai les »), à l'exception de l'article. Lorsque ce n'est pas le cas, il s'agit d'un faux-départ ;

- Les faux-départs : ils se distinguent des autocorrections par le fait que :

**« contrairement à l'autocorrection, il n'existe aucune analogie entre le segment remplacé et le reste de l'énoncé. » (Kurdi, *op. cit.*, p. 25)**

Pour illustrer cette définition, l'auteur emploie l'exemple suivant (là encore, la disfluenza est soulignée) :

(...) oui c'est à e ça se prend au deuxième étage

Dans cet exemple, il n'y a pas d'analogie entre le segment remplacé (« c'est à ») et sa correction (« ça se prend ») ;

- Les incomplétudes : ce terme s'applique aux énoncés (et non aux mots) qui ne sont pas terminés, autrement dit qui sont incomplets syntaxiquement. Comme l'auteur le remarque, ce phénomène n'a pas été pris en compte par les principales études sur les disfluences.

#### 4. Conclusion

Ce chapitre nous a permis de faire une première présentation des phénomènes qui font l'objet de cette thèse. Nous avons montré que leur apparition est naturelle dans le cadre de toute production orale spontanée, qu'il s'agisse de discours ou de dialogues.

Nous avons également fait une synthèse des principales études sur le sujet. Ces études ont montré que, malgré l'apparence aléatoire revêtue par les disfluences, leurs manifestations obéissent en fait à certaines régularités, notamment d'ordre syntaxique. Cette synthèse nous a également permis de considérer les causes des disfluences et plus généralement des problèmes apparaissant à l'oral spontané, d'un point de vue cognitif et psycholinguistique. Enfin, notre but est de pouvoir ainsi disposer de catégorisations des disfluences, qui sont un phénomène particulièrement difficile à classer en raison de leur grande variété de manifestations. Il est cependant possible de dégager quelques grandes catégories, même si les frontières entre elles sont assez floues et varient selon les auteurs.

A partir de ce panorama des catégorisations (dont nous faisons la synthèse<sup>19</sup> dans le Tableau 1 ci-dessous), nous pourrions choisir la terminologie qui est la plus adaptée au contexte que nous étudions.

**Tableau 1** : Présentation synoptique des catégorisations de disfluences

Auteurs	Shriberg	Candea	Henry et Pallaud	Kurdi
Nature du corpus	3 corpus avec des situations variées de dialogues oraux spontanés	Histoires racontées oralement par des enfants	<i>Corpaix</i> : dialogues oraux spontanés (différentes situations)	Négociations de transport de marchandises
Phénomènes étudiés/terminologie employée	<ul style="list-style-type: none"> <li>• Répétitions</li> <li>• Pausés pleines</li> <li>• Termes explicites d'édition</li> <li>• Marqueurs du discours</li> <li>• Fragments de mots</li> <li>• Insertions</li> <li>• Suppressions</li> <li>• Substitutions</li> </ul>	<ul style="list-style-type: none"> <li>• Pausés silencieuses</li> <li>• « euh » dits d'hésitation</li> <li>• Allongements vocaliques</li> <li>• Répétitions</li> <li>• Autocorrections</li> </ul>	<ul style="list-style-type: none"> <li>• Répétitions</li> <li>• Pausés</li> <li>• Fragments de mots (ou amorces) : <ul style="list-style-type: none"> <li>○ Inachevés</li> <li>○ Complétés</li> <li>○ Modifiés</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Extragrammaticalités lexicales : <ul style="list-style-type: none"> <li>○ Pausés</li> <li>○ Mots oraux</li> <li>○ Mots incomplets, amalgames.</li> </ul> </li> <li>• Extragrammaticalités Supralexicales : <ul style="list-style-type: none"> <li>○ Répétitions</li> <li>○ Auto-corrections</li> <li>○ Faux-départ</li> <li>○ Incomplétudes.</li> </ul> </li> </ul>

Cependant, les corpus sur lesquels les études que nous avons présentées ont été menées peuvent être qualifiés de « tout venants ». Par ce terme, nous voulons dire que leurs caractéristiques linguistiques relèvent du langage (français ou anglais) tel qu'il est employé dans la majorité des situations de la vie quotidienne. Or, le travail décrit dans notre thèse porte sur un corpus très différent à tout point de vue. Les différences sont telles qu'il est nécessaire de leur consacrer l'intégralité du prochain chapitre.

<sup>19</sup> A l'exception du travail de C. Blanche-Benveniste : comme nous l'avons vu en 3.2.2, il n'y a pas à notre connaissance de travail de classification systématique de disfluence entrepris.

## Chapitre 2

# L'oral spontané dans le contrôle de la navigation aérienne

### 1. Introduction

Dans le premier chapitre, nous avons passé en revue les principales définitions des phénomènes de disfluences, et décrit comment elles se manifestent au sein de divers corpus (réservations, conversations, etc.) de dialogues oraux spontanés.

Le travail que nous rapportons porte sur des corpus différents de ceux présentés dans le chapitre 1. Il s'agit de dialogues relatifs au contrôle aérien. La tâche de contrôle aérien entraîne des contraintes particulières qui s'exercent notamment aux niveaux linguistique et cognitif. Nous montrons dans le présent chapitre que ces contraintes sont bien plus fortes que celles mises en œuvre ou rencontrées dans les dialogues relatifs à d'autres tâches. Nous répondrons ainsi à l'une des questions soulevées dans le chapitre précédent : ces contraintes peuvent-elles avoir une influence sur la nature, le mode de manifestation, et la distribution des phénomènes de disfluences ?

Dans un premier temps, nous décrirons un aperçu général du contexte dans lequel s'inscrit le contrôle aérien. Il s'agit d'un domaine très spécialisé et peu connu en dehors des milieux qui le pratiquent. Nous affinerons ensuite la granularité de notre analyse en nous penchant sur les caractéristiques des dialogues entre contrôleurs aériens et pilotes. Enfin, nous présenterons les corpus d'étude, en les plaçant dans la perspective définie dans les sections précédentes.

### 2. Présentation générale du contexte

#### 2.1. Le trafic aérien

Le trafic aérien consiste en la circulation de l'ensemble des différents appareils volant dans un périmètre donné (une région, un pays, etc.). Cette apparente banalité recouvre une très grande complexité : dans la zone aérienne d'un pays tel que la France, des milliers d'avions peuvent évoluer au même moment, qu'il s'agisse de vols civils ou militaires. De plus, pour des raisons de sécurité, chaque avion doit être suffisamment distant de tout autre appareil volant pour éviter, non seulement toute collision, mais également tout phénomène pouvant être causé par une interaction physique trop proche, tel que des appels d'air par

exemple. La situation est encore plus complexe en raison de la croissance permanente du nombre de vols affrétés<sup>20</sup>. En effet, le nombre d'avions à traiter dans une zone donnée augmente ; de plus, comme chaque avion doit disposer d'un certain espace pour se mouvoir sans risque de collision, les plans de vols correspondants sont d'autant plus difficiles à établir.

Il serait impossible de faire fonctionner un tel système complexe de manière ordonnée sans une série de mécanismes de gestion et de contrôle, au centre desquels se trouvent les contrôleurs aériens, et que nous décrivons ci-après.

## **2.2. La tâche de contrôle aérien**

D'une manière générale, la tâche de contrôle aérien vise à remplir deux buts : informer d'une part, et guider d'autre part, les différents véhicules volants<sup>21</sup>. Tous les centres de contrôle aérien ne se situent pas en proximité des aérodromes. Il y a cinq centres appelés « Centres en Route de la Navigation Aérienne » qui servent à gérer le trafic entre les aérodromes ; il y a également les différents centres situés eux, à proximité des aérodromes, et permettant de contrôler l'approche et l'atterrissage sur l'aérodrome. De fait, il est exprimé en aéronautique une distinction entre les deux aspects du vol : *en route* d'un côté et *en approche* de l'autre. Ils correspondent, en effet, à deux types de procédures assez différentes, que ce soit au niveau des résultats obtenus ou des moyens mis en œuvre. Ainsi, la procédure d'*approche* consiste essentiellement à gérer l'arrivée à l'aéroport, l'atterrissage, etc., alors que celle d'*en route* implique notamment la gestion du plan de vol (les différentes positions géographiques par lesquelles doit passer l'aéronef). Celles-ci ont également des points communs : par exemple, la gestion de la cohabitation de différents vols, la variabilité des conditions météorologiques, etc.

Dans le cadre de notre recherche, nous étudions les dialogues opérationnels dans les Centres en Route. Il n'y a fondamentalement pas de différence entre les tâches des contrôleurs d'un ou de l'autre type de centre, surtout en ce qui concerne la *phraséologie*. Ce terme désigne le langage employé par les opérateurs humains impliqués dans le trafic aérien. Il s'agit d'un ensemble lexico-syntaxique (lexique, règles de composition des énoncés) constitué par les différents ordres et informations relatifs au contrôle aérien. Nous présentons les caractéristiques de ce langage dans la section 3 ci-dessous.

Le travail du contrôleur aérien est de remplir deux objectifs. Le premier est la communication d'information aux pilotes, de manière à leur permettre de naviguer dans les meilleures conditions possibles de sécurité. Ces informations concernent les conditions météorologiques, les changements de cap d'autres aéronefs, etc.

<sup>20</sup> Cette croissance est due à des raisons économiques et sociales qui ne sont pas à décrire ici.

<sup>21</sup> Il est à noter qu'il existe également dans le cadre du contrôle aérien français des agents chargés de la seule information des pilotes, indépendamment de leur guidage. Mais dans le cadre de cette thèse, on ne traitera que le cas des seuls contrôleurs au sens propre du terme.

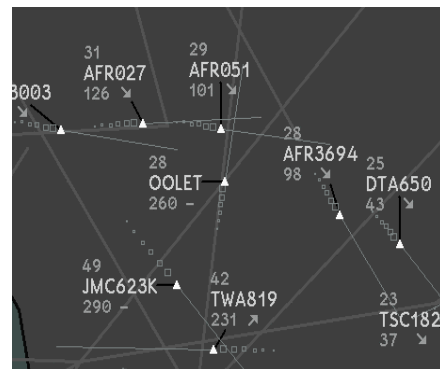
Le deuxième objectif est le contrôle aérien. Il consiste principalement à faire respecter le plan de vol défini à l'avance pour les avions, à en gérer les éventuelles modifications, sources de conséquences sur les autres plans de vol, etc. Pour cela, les contrôleurs ont recours à des instructions, appelées *clairances*, donnant des indications sur les actions à effectuer compte tenu de la situation courante et du plan de vol. Ces indications peuvent avoir trait à l'attitude de l'avion (maintenir, monter, baisser), sa vitesse, sa direction, etc.

Le but principal du contrôle aérien est d'assurer la sécurité des différents vols. Les dangers à éviter sont toute collision ou problème imprévu lié à la cohabitation de différents avions dans le même secteur, ainsi que tout autre perturbation, qu'elle soit météorologique, humaine, mécanique, etc. Chaque aéronef est identifié par un indicatif alphanumérique (par exemple « D T C » ou « Britair 452 »), et doit suivre un plan de vol indiqué par le contrôleur, en fonction des trajectoires des autres avions. Chaque contrôleur s'occupe d'un secteur particulier.

Les contrôleurs travaillent généralement en binôme (cf. Figure 5). Leurs activités de contrôle s'appuient sur deux sources d'informations : la première, numérique, est supportée par l'affichage d'informations sur les avions du secteur : un écran radar donnant les indications sur leur position, leur vitesse, etc. (Figure 6). La seconde est donnée par des « *strips* » (Figure 7). Ces derniers sont des bandes de papier sur lesquelles les contrôleurs notent diverses informations relatives aux avions de leur zone, au fur et à mesure des déplacements de ceux-ci dans la zone contrôlée. Les *strips* servent à la fois d'aide-mémoire, d'archive, ainsi que de support pour la transmission des données aux autres contrôleurs, lorsque les appareils passent dans leur secteur.



**Figure 5** : Contrôleurs aériens en exercice



**Figure 6** : Exemple d'image radar



**Figure 7** : Exemple d'un strip

La tâche des contrôleurs aériens est assistée par des outils informatiques. Nous pensons notamment au *Data-Link*. Il s'agit d'un dispositif technique qui permet la transmission de données air-sol par liaison<sup>22</sup> de données numériques. Le *Data-Link* présente de nombreux intérêts, parmi lesquels la possibilité de diffuser des informations complémentaires à la communication orale. De plus, de nombreuses recherches sont menées afin d'optimiser les différents types d'interaction, voire en proposer de nouveaux<sup>23</sup>.

Cependant, pour l'instant, la principale modalité d'interaction utilisée par les contrôleurs reste la communication orale avec les autres acteurs du trafic aérien, en particulier les pilotes. Cette modalité présente un certain nombre de caractéristiques spécifiques au trafic aérien, qui font l'objet du développement suivant.

### **3. Le dialogue oral dans le contrôle aérien**

Dans le vaste domaine que recouvre le contrôle aérien, la communication orale est utilisée non seulement entre les contrôleurs et les pilotes, mais également entre les contrôleurs entre eux, les pilotes entre eux, etc.

Les motivations de l'étude du seul dialogue contrôleur-pilote (pour l'utilisation de la modalité orale) sont d'ordre linguistique.

En effet, l'étude des dialogues *contrôleur-pilote* est intéressante car les deux locuteurs sont placés dans des conditions pragmatiques différentes : au niveau de la situation dans laquelle se trouvent les locuteurs (par exemple, piloter un avion en déplacement implique des contraintes matérielles différentes que de se trouver dans une tour de contrôle) et au niveau fonctionnel (les rôles sont très différents et bien codifiés). Par contre, le contenu linguistique des différentes situations est à peu près identique : les locuteurs, quel que soit leur rôle dans le trafic aérien, utilisent la même phraséologie.

Nous débutons cette section par une présentation générale des principales caractéristiques des dialogues contrôleurs / pilotes. Nous continuons ensuite cette analyse, de manière plus approfondie, en nous basant sur quatre études menées sur la phraséologie du contrôle aérien : Falzon (1982 et 1989), et Volpe *et al.* (2002 et 2004)<sup>24</sup>. Notre but sera de montrer en quoi, et comment, les caractéristiques pragmatiques dictées par la tâche (particulièrement les contraintes de sécurité du trafic aérien) influent sur les aspects linguistiques.

---

<sup>22</sup> Liaison de type radio-électrique. Les données transmises par ce biais sont numériques et non analogiques comme cela est le cas avec les communications classiques.

<sup>23</sup> Voir notamment les travaux menés par l'équipe *PII* du DSNA (<http://www.tls.cena.fr/divisions/PII>), par exemple Mertz *et al.* (2000), Mertz (2003).

<sup>24</sup> Signalons que les études menées sur ces aspects de la communication de contrôle aérien sont très rares. A notre connaissance, celles que nous citons ici sont les seules existantes, au moins en ce qui concerne la sphère francophone.

### 3.1. Les dialogues contrôleurs aériens / pilotes

Comme nous le mentionnons dans le chapitre 1 (section 2.2), tout dialogue doit respecter des règles, implicites ou explicites, et être structuré. Cet impératif est primordial dans le cas des dialogues de contrôle aérien. Dans un premier temps, nous expliquons les raisons pour lesquelles ce type de dialogue nécessite l'existence de la phraséologie, que nous avons définie dans le présent chapitre, en section 2.2. Dans un second temps, nous présentons les caractéristiques de celles-ci.

#### 3.1.1. Impératifs de la communication

La communication des contrôleurs et pilotes est totalement dictée par la tâche de contrôle (voir section 1.2 ci-dessous). Les principaux impératifs à respecter pour la communication dans le cadre du contrôle aérien sont :

- L'intelligibilité : il faut prendre en compte les problèmes de transmission radio ;
- La rapidité : le contrôle aérien s'effectue dans un contexte particulièrement dynamique : les avions en circulation se déplacent très rapidement ; des imprévus peuvent arriver (pannes, changements météorologiques, etc.) et demander une réaction instantanée. Il est, par conséquent, impératif que les communications relatives à la tâche de contrôle soient aussi rapides que possible ;
- La bonne interprétation du message : il est nécessaire d'être certain que le message transmis a bien été bien réceptionné et interprété. Un des points cruciaux est la suppression de toute ambiguïté. Cette nécessité concerne le niveau sémantique : à un mot ou une instruction donné ne doit correspondre qu'un seul sens, strictement défini. De ce point de vue, la phraséologie est l'exacte représentation de la portion de « monde réel » constitué par le trafic aérien, sans ajout ni retrait d'information supplémentaire. Pour reprendre la terminologie de Jakobson (1963), les fonctions référentielles et conatives sont primordiales, la fonction phatique un peu moins, et les fonctions métalinguistique, expressive et poétique inexistantes.

#### 3.1.2. Caractéristiques de la phraséologie du contrôle aérien

Pour répondre à ces impératifs, la communication des contrôleurs est dictée par une phraséologie, décrite officiellement dans (*Arrêté du Journal Officiel*, 2000<sup>25</sup>). Le contrôleur aérien doit maîtriser cette phraséologie. L'objectif de la phraséologie est décrit dès les premières pages du document officiel que nous rappelons brièvement ci-dessous :

---

<sup>25</sup> Légèrement modifié par l'Arrêté du 24 novembre 2005, paru dans le *J.O* n° 26 du 31 janvier 2006, page 1621. Téléchargeable à [www.sia.aviation-civile.gouv.fr/dossier/texteregle/RDA\\_TA\\_GEN\\_091.pdf](http://www.sia.aviation-civile.gouv.fr/dossier/texteregle/RDA_TA_GEN_091.pdf)

- L'étude d'incidents et d'accidents a mis en évidence, en tant que facteurs contributifs, le non-respect des procédures de radiotéléphonie ou l'emploi d'une phraséologie approximative ;
- C'est pourquoi, lors de communications radiotéléphoniques, les pilotes et le personnel des stations au sol doivent respecter les procédures de radiotéléphonie et, dans la mesure du possible, utiliser des expressions conventionnelles et une phraséologie normalisée (*Arrêté du Journal Officiel*, 2000, p. 9) ;

La phraséologie décrit l'ensemble des mots et des règles que doit respecter la production d'un message : par exemple, sa priorité (un message prioritaire doit être transmis avant les autres)<sup>26</sup>, les différentes expressions à utiliser, l'ordre que doivent respecter les différents constituants de l'énoncé. Elle doit couvrir de manière exhaustive l'ensemble des situations auxquelles peuvent être confrontés des opérateurs du trafic aérien.

Les ordres peuvent être *simples ou complexes*. Un ordre simple correspond à une seule instruction donnée par le contrôleur. Un exemple d'un ordre simple qu'un contrôleur aérien peut donner à un pilote est

*Indicatif      Ordre*

Cet ordre se décompose en deux sous-ensembles syntagmatiques : l'indicatif de l'avion (« D T C »), et ensuite l'ordre lui-même.

Les ordres complexes sont constitués de la combinaison de plusieurs commandes simples. Par exemple, l'énoncé

*Indicatif      Politesse      Premier ordre      Deuxième ordre*

est constitué respectivement de l'indicatif de l'avion auquel s'adresse le contrôleur, de la formule de politesse par lequel le contrôleur indique dans quel centre il opère, et des instructions « maintenez niveau 180 route Poitiers Amboise Lacan » et « rappelez Amboise ». La première instruction indique au pilote du vol F G H M I qu'il doit se maintenir au niveau 180 (l'unité est le pied (« *foot* » en anglais, abrégé *ft.*) en suivant une route définie par trois centres. Dans la deuxième instruction, il est demandé au pilote de faire un compte rendu lorsqu'il se trouvera au niveau d'Amboise.

<sup>26</sup> La priorité est définie par une hiérarchie : un message de détresse passe avant tout les autres, viennent ensuite les messages d'urgence, puis de contrôle du trafic, puis d'information, et enfin la communication entre exploitants d'aéronefs.



Jusqu'ici, nous avons considéré les énoncés dans leur individualité, c'est-à-dire en tant que tours de parole distincts les uns des autres. Voici à présent deux exemples de dialogues entre des contrôleurs et des pilotes, issus de situations réelles.

1<sup>er</sup> dialogue (2 tours de parole) :

**Contrôleur** : F P I contacter Paris 120,95 au revoir.

**Pilote** : Paris 120,95 PI au revoir.

Commentaires : Voici un exemple d'un dialogue très simple, puisqu'il ne comporte qu'un seul échange entre le contrôleur et le pilote, chaque échange ne comportant qu'un seul énoncé. Le contrôleur s'adresse au pilote du vol F P I (il le mentionne explicitement pour éviter toute ambiguïté), et lui enjoint de contacter le centre de Paris, en utilisant la fréquence « 120,95 »<sup>27</sup> du canal. Le pilote collationne en reprenant les éléments de l'instruction du pilote. On remarque que le pilote utilise plusieurs « raccourcis » : omission du verbe « contacter », et forme abrégée de l'indicatif<sup>28</sup>. On voit ici que la phraséologie, malgré le cadre contraint qu'elle fixe, permet néanmoins une certaine souplesse de l'expression.

2<sup>ème</sup> dialogue (9 tours de parole) :

**Contrôleur** : COTAM 12 75 bonjour.

**Pilote** : Bonjour.

**Pilote** : 12 75 au niveau 90 en route vers SO SOPIL.

**Contrôleur** : Roger vous choisissez un vous voulez un niveau supérieur ?

**Pilote** : Négatif on doit rentrer rapidement c'est pour une Evasan.

**Contrôleur** : Alors vous montez au niveau 100 ou vous choisissez 100 ou 80 comme vous voulez.

**Pilote** : 80.

**Contrôleur** : Alors euh descendez au niveau 80 et procédez sur AMB.

**Pilote** : AMB merci.

Commentaires : nous avons choisi ce dialogue pour donner un exemple de discussion plus complexe que la précédente entre le contrôleur et le pilote ; on y voit que le dialogue du contrôle aérien peut être plus complexe que l'exemple précédent. La situation considérée ici permet de voir comment les deux dialoguants s'adaptent à une situation donnée : le pilote est dans une situation d'urgence (une *Evasan* désigne une *EVAcuation*

---

<sup>27</sup> Rappelons ici qu'un numéro de fréquence correspond à la zone d'une plage de longueur d'onde (le canal) pouvant être utilisée pour une communication donnée.

<sup>28</sup> Selon les circonstances, l'indicatif désignant un avion peut être prononcé selon différentes formes, de longueurs variables. Par exemple, à la forme complète de l'indicatif « FPI 215 » pourra être substitué « P I », comme dans le dialogue que nous considérons ici, ou « 2 1 5 », etc. Les différentes dérivations pouvant être obtenues à partir d'un indicatif donné résultent de l'application de règles. Ces règles ont été décrites et formalisées dans (Dourmap, 2003).

*SANitaire*), et le contrôleur lui laisse une certaine latitude par rapport au choix de l'orientation du vol.

La phraséologie cherche également à supprimer les ambiguïtés sur le plan phonétique : les éléments pouvant faire l'objet d'une confusion doivent être articulés clairement. Par exemple, toute lettre ou série de lettres isolées (constituant, par exemple, des indicatifs) doit être épelée en respectant l'alphabet utilisée dans la radio-émission : « Alpha » pour « A », « Bravo » pour « B », « Charlie » pour « C », etc.

### **3.2. Analyses linguistiques des dialogues contrôleurs-pilotes**

Les caractéristiques décrites ci-après sont tirées des quatre principales études. Pour bien en appréhender la valeur scientifique, en voici d'abord une présentation générale :

- Falzon (1982 et 1989) : Falzon (1982) est fondé sur l'hypothèse suivante  
**« le langage des contrôleurs ne doit pas être réduit à un langage de commande, il n'est pas non plus un échantillon de la langue naturelle : il en est une restriction. »**  
**(Falzon (1982, p.1))**

L'ensemble de l'étude est consacrée à l'évaluation de cette hypothèse, en prenant en compte tous les paramètres linguistiques disponibles. Falzon (1989) reprend le contenu de Falzon (1982), et apporte en plus différentes analyses d'ordre cognitif ;

- Volpe et al (2002 et 2004) : le but de ces études est d'analyser la manière dont le canal de communication air-sol est utilisé. L'analyse comprend aussi bien un volet structurel, relatif au taux d'occupation du canal, au nombre de communications passées, etc. qu'un aspect plus sémantique, avec étude du contenu des dialogues. Les deux études sont menées respectivement à partir de conversations *en route* et *en approche*. A cette différence près, les objectifs et méthodologies sont identiques.

Des analyses de la phraséologie faites dans ces études, il ressort que le lexique est finalisé à l'application et que les structures syntaxiques de l'oral sont simplifiées. Il est également constaté que certains mots et fréquences apparaissent beaucoup plus fréquemment que d'autres. Revenons sur ces différents points :

- Lexique finalisé à l'application : les deux caractéristiques du lexique utilisé par la phraséologie sont d'une part, une taille réduite (mesurée en nombre d'éléments lexicaux) et d'autre part, une grande adaptation à la spécificité du contrôle aérien.
  - Taille du lexique : Falzon (1982, p. 13) dénombre 417 mots différents en français et 214 mots différents en anglais, soit un total de 714 mots différents

utilisés. Pour rappel, la taille du lexique français « fondamental »<sup>29</sup> est d'approximativement 4000 à 5000 mots<sup>30</sup> ;

- Distribution des classes de mots du lexique du contrôle aérien : Falzon (1982, p. 18-19) fait également une catégorisation fine des messages transmis sur le canal de communication, et de leur fréquence, en sept classes de messages différents. Trois de ces classes sont exclusives au domaine du contrôle aérien (gestion du trafic aérien, du contact radio, et du radar) ; les 4 autres ne sont pas aussi spécifiques : politesse, coordination des messages (« et » ...), « méta-messages » (« dommage », « tant mieux »), et mise à jour des connaissances. Leur distribution indique une forte adaptation à la tâche : les seules classes relatives à la gestion du contrôle aérien et celle du contact radio couvrent respectivement près de 40% et près de 30% de l'ensemble des messages prononcés (Falzon (1982, p. 25)).
- Fréquences d'emploi : Parmi les différentes conclusions de Falzon, nous en voyons deux qui sont particulièrement intéressantes dans le cadre d'une analyse linguistique de la phraséologie. La première concerne le taux de couverture lexicale : **« pour couvrir 90% des mots prononcés, il suffit de 130 mots différents en français (c'est-à-dire 27.4% du vocabulaire français), 84 mots différents en anglais (c'est-à-dire 34% du vocabulaire anglais). »** (Falzon, op. cit., p. 15). Cette constatation est intéressante car elle permet de voir que non seulement le lexique correspondant à la phraséologie est très restreint, mais que de plus l'utilisation effective des mots qui en est faite est encore plus limitée ! Cela peut paraître contradictoire avec une autre observation de Falzon, concernant ce qui est appelé la *banalité* des mots. La définition en est donnée p.15 (voir aussi p. 28) :  
**« l'indice de banalité d'un mot est le nombre de contrôleurs qui utilisent ce mot. »** (Falzon, op. cit.).  
Il s'avère, d'après les différentes statistiques effectuées, que les mots d'indice 19 à 20 (maximum de la notation) sont très peu nombreux (3 en français, 5 en anglais). Autrement dit, les contrôleurs ont propension à utiliser chacun des mots différents de ceux de leurs collègues. Ce résultat peut sembler étonnant. En effet, la synonymie est rendue impossible par la nécessité absolue d'éviter toute ambiguïté. D'autre part, la taille réduite du lexique (plusieurs fois inférieure à celle du français « fondamental ») restreint le choix parmi les différents lexèmes pouvant être utilisés pour exprimer une idée donnée. Une des explications à cette stratégie des locuteurs

---

<sup>29</sup> Notion évidemment assez floue et intuitive, utilisée entre autres par l'Académie Française, désignant l'utilisation au quotidien de la langue française.

<sup>30</sup> Voir par exemple le travail de Catach *et al.* (1984), établissant une liste des 1600 mots les plus fréquents du français, à partir de laquelle est défini un lexique de base du français.

est le recours à la langue anglaise pour exprimer certaines instructions de changement de niveau ; cela conduit à une diminution de la banalité.

D'un point de vue linguistique, le langage utilisé par les contrôleurs est « une restriction de la langue naturelle » (voir Falzon, 1989, chapitre 6). J.-M. Pierrel introduit, lui, la notion de « langage naturel finalisé » (cf. par exemple (Pierrel, 2000)). Comme ces termes l'indiquent, il s'agit d'un langage naturel (par opposition, par exemple, aux langages de programmation ou de commande), mais dont les caractéristiques sont différentes de l'ensemble auquel il appartient (par exemple, le français). Nous reviendrons plus en détail sur ce point dans la section 3.5.2.

La phraséologie ne prévoit pas l'utilisation de procédés ordinairement employés à l'oral : incises, répétitions « stylistiques » qui ne sont pas des disfluences à proprement parler mais participent à la difficulté de traitement de la syntaxe orale. De ce fait, il est prévisible de ne pas en trouver beaucoup d'occurrences.

De fait, cette phraséologie n'est pas toujours strictement respectée, même si les recommandations générales qu'elle donne sont appliquées. Le fait que les communications soient régies offre un cadre d'étude privilégié pour étudier les déviations par rapport à la phraséologie.

L'acquisition de celle-ci nécessite une phase d'apprentissage très stricte. La prochaine sous-section lui est consacrée.

### ***3.3. Les dialogues d'apprentissage et d'évaluation***

L'apprentissage de la phraséologie est une étape cruciale de la formation des futurs contrôleurs aériens. Sur le plan linguistique et cognitif, la situation d'apprentissage induit certaines particularités par rapport à une situation « réelle », que nous présentons dans la section 3.5. De plus, dans la problématique qui nous intéresse – la production d'erreurs et les stratégies de correction – il s'agit d'un cadre privilégié d'étude. En effet, on peut s'attendre à ce que ces phénomènes y soient plus fréquents étant donné le manque d'expérience des locuteurs.

La formation des contrôleurs aériens se fait à l'*Ecole Nationale d'Aviation Civile (ENAC)*, dont un centre se trouve à Toulouse. Tout au long de leur cursus, les contrôleurs effectuent divers exercices d'entraînement et d'évaluation. L'un de ces exercices consiste en sessions de dialogues oraux spontanés entre des contrôleurs aériens en formation et des *pseudos-pilotes* (cf. Figure 8). Ces derniers sont des instructeurs simulant le rôle de pilotes en vol, mais situés dans les locaux de l'ENAC.

Les sessions d'apprentissage sont conçues de manière à simuler autant que possible des conditions réelles de travail. Ainsi, les communications sont faites via le canal HF<sup>31</sup>, les contrôleurs s'exercent devant un radar donnant des positions d'avions fictifs, pourvus chacun d'un plan de vol lui aussi fictif, etc. Comme dans une tâche de contrôle réelle, les élèves contrôleurs doivent gérer plusieurs avions situés dans la zone contrôlée (appelée *secteur*), par exemple, en leur assignant une vitesse et/ou une position données. Pour cela, ils interagissent oralement avec les pseudos-pilotes selon les prescriptions de la phraséologie définie ci-dessus (cf. 3.1.2).

Cependant pour ceux-ci, la situation et l'environnement matériel sont différents. En effet, les « pseudo-pilotes » ont accès à une interface graphique leur permettant de visualiser les mêmes données radar que celle visionnées par le contrôleur, mais également de saisir (au moyen d'un clavier et d'une souris) les instructions données par ce dernier, afin que l'avion virtuel se comporte conformément aux instructions : changement de niveau, de cap, etc. Les pseudo-pilotes disposent également d'un dispositif de communication orale avec les élèves contrôleur (qui lui-même est équipé d'un système audio similaire). Ce dispositif est constitué d'un microphone et d'un bouton-poussoir<sup>32</sup> permettant de déclencher et stopper la diffusion de la voix du locuteur sur le canal.

Pour faciliter la tâche du « pseudo-pilote », différentes études ergonomiques de l'interface sont menées afin d'en améliorer le confort d'utilisation et la rapidité d'exécution (voir par exemple Drut *et al.* (2008)).

L'ensemble du processus d'interaction Pseudo-pilote / Contrôleur et des dispositifs afférents est schématisé dans la Figure 8 ci-après (empruntée à Schäfer (2001, p. 17)).

---

<sup>31</sup> *Canal Haute Fréquence*. Ce terme technique désigne l'utilisation d'ondes courtes, dont une des particularités est la possibilité de réflexion sur des couches de l'atmosphère. De ce fait, elles sont particulièrement adaptées à la transmission de communications à moyenne et longue distance, comme cela est souvent le cas dans l'aéronautique.

<sup>32</sup> Ordinairement une pédale.

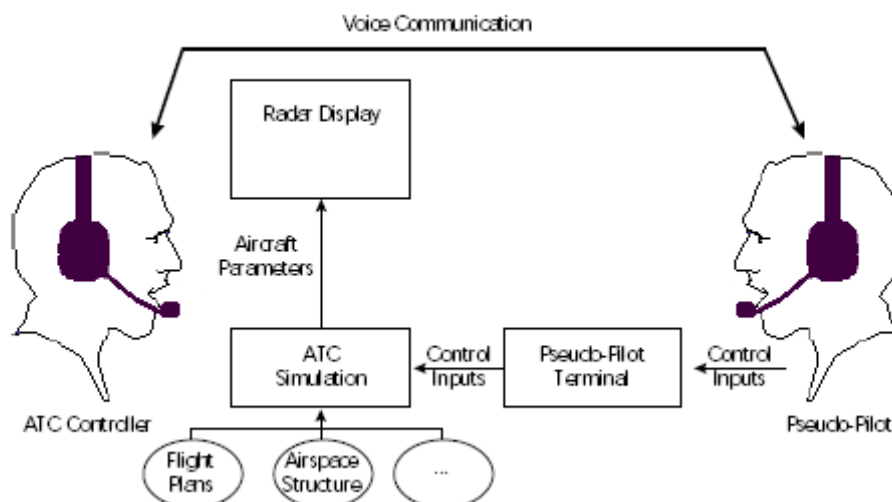


Figure 8 : Communication pseudo-pilote / Contrôleur

### 3.4. La prévention des erreurs humaines dans le contrôle aérien

Jusqu'à présent, nous avons montré que l'organisation de la tâche de contrôle aérien a été conçue de manière à prendre en compte les différents paramètres du contexte aéronautique. Un autre paramètre pris en considération est la faillibilité de la communication.

Celle-ci est particulièrement critique, comme le montrent de nombreuses études (voir notamment Davison (2003) pour les problèmes liés à la communication dans l'aéronautique, ou Dehais (2004) pour l'impact des processus cognitifs, notamment la *dissonance cognitive*<sup>33</sup>, sur le pilotage). Cette prise en compte se manifeste de différentes manières selon le rôle de l'opérateur considéré (contrôleur, pilote), la tâche qu'il accomplit (pour un contrôleur : communication avec le pilote, avec les autres contrôleurs, consultation des plans de vol ... ; pour le pilote : communication avec le contrôleur, gestion de l'avion ...), etc.

Dans le dialogue oral de contrôle aérien, il est prévu des mécanismes de prévention, ou au moins réduction, des risques d'erreurs. Ils permettent, en effet, aux locuteurs de s'adapter aux éventuelles difficultés de communication, en demandant de parler plus lentement, de répéter, d'interroger sur la qualité de la réception audio<sup>34</sup>, etc. A chacune de ces actions correspond une expression définie par la phraséologie : « comment recevez-

<sup>33</sup> « Un individu se trouve dans un état de dissonance cognitive lorsqu'il est dans des situations non cohérentes, comme par exemple réagir contrairement à son point de vue » (Dehais, *op. cit.* p. 56). Dans le contexte du trafic aérien, cela peut se traduire par des comportements dangereux à terme ; par exemple, persévérance dans un plan de vol fixé à l'avance, malgré des conditions telles que le manque de carburant, une tempête, etc.

<sup>34</sup> Par exemple, « me recevez-vous », « reçu 5 sur 5 ».

vous », « confirmez », « parlez plus lentement », etc.<sup>35</sup> (cf. Arrêté du Journal Officiel (2000, p.15-18)).

En plus de ces expressions, le principal dispositif à l'œuvre est le système de *collationnement*. Il consiste en la répétition, par le pilote, de tous les ordres qu'il vient de recevoir de la part du contrôleur. Ainsi, ce dernier peut vérifier immédiatement si ses instructions ont été correctement entendues et comprises. Le contrôleur peut également collationner les messages du pilote lorsque cela est nécessaire, notamment quand le message du pilote comporte des éléments qui seront utilisés par le contrôleur pour donner la *clairance* (rappelons que ce terme est synonyme d'ordre ou d'instruction). Le concept de collationnement peut sembler en opposition avec la nécessité, évoquée plus haut, de rapidité de la communication. Mais il correspond totalement au besoin de précision et de sûreté. Il s'agit par conséquent d'un compromis équilibré entre les deux impératifs. Il est, d'ailleurs, employé dans certains autres contextes nécessitant eux aussi un temps de réaction très rapide, par exemple lors d'opérations militaires.

Même s'il permet d'éviter un grand nombre de problèmes, le principe n'est toutefois pas parfait. En effet, il est, par exemple, possible que le contrôleur ne soit pas attentif à une erreur collationnée par le pilote. Par conséquent, la nécessité de rendre aussi robuste que possible leur communication reste intacte. D'autant plus que la réaction de l'opérateur humain, s'il se rend compte d'un problème, peut elle-même induire d'autres problèmes. Par exemple, la correction d'une erreur, bien que nécessaire, introduit des ambiguïtés du fait de la juxtaposition de l'élément erroné et de sa correction, comme on peut le voir dans l'exemple suivant :

Air Vendée 44 78 **contactez correction restez** sur la fréquence je vous rappelle

Nous revenons amplement sur ces points dans les chapitres 3 et 4, grâce à une analyse fine de corpus.

### **3.5. Conséquences et caractéristiques cognitives et linguistiques**

Comme dans l'ensemble de cette thèse, nous envisageons les conséquences des propriétés des dialogues de contrôle aérien, que nous venons de décrire, selon une approche pluri-disciplinaire. Nous nous plaçons dans un premier temps dans une perspective cognitive, puis linguistique dans un deuxième temps.

#### **3.5.1. Caractéristiques cognitives**

L'une des causes de disfluences ou d'erreurs chez les sujets sains est la présence d'une charge de travail (ou charge cognitive)<sup>36</sup> trop élevée et/ou de conditions environnementales

---

<sup>35</sup> Chacune de ces expressions a un équivalent en langue anglaise, également défini.

<sup>36</sup> Ce terme désigne la mobilisation des ressources cognitives de l'individu : attention, mémoire de travail, etc. (cf. Spérando (1984)).

bruitées (Scherer, 1986). Lorsque cela se produit, le traitement par le locuteur de sa propre production linguistique est rendu plus difficile<sup>37</sup>. Comme nous l'avons montré dans le premier chapitre, ce traitement comporte plusieurs phases allant de la planification à la réalisation, avec une auto-évaluation du résultat produit par rapport à l'intention de départ. Chacune de ces phases peut être perturbée par des conditions difficiles de production.

Cette situation se rapproche alors de celle d'un locuteur atteint de troubles pathologiques (bégaiement, absence de mots, etc. cf. Nespoulous (1990 ; 1996 ; 2004a et b)).

Comment est modifiée la charge cognitive ? Les études sur les facteurs humains dans le contrôle aérien (notamment Athènes (2002), Averty (2002 ; 2004)) rapportent que les principaux paramètres affectant la charge cognitive des contrôleurs aériens sont le nombre d'avions à gérer à un instant  $t$ , et la limitation du temps imparti pour donner des ordres.

De même que la charge cognitive, la situation d'apprentissage joue un rôle sur les performances de production vocale des sujets enregistrés. Nous formulons l'hypothèse que les contrôleurs en formation ne se sont pas encore suffisamment approprié les tâches de gestion des avions et la phraséologie. Pour un même nombre d'avions à gérer, la charge de travail de l'élève contrôleur serait, par conséquent, augmentée par rapport à celle d'un contrôleur expérimenté. Cette hypothèse paraît partagée par (Averty *et al.*, 2002), qui écrivent (en parlant de l'activité du contrôleur aérien) :

**“Moreover, repetitive and practical experience of these strategies in a given, well-known airspace results in a highly automated set of solutions, requiring probably few cognitive resources.” (Averty *et al.*, 2002, p. 2)**

Or, selon notre hypothèse, cette augmentation a pour corollaire une plus grande difficulté pour mener à bien le processus de réalisation linguistique. Cette difficulté est accrue par la non-maîtrise de la phraséologie : le locuteur ne s'étant pas encore suffisamment approprié le lexique, ni les règles syntaxiques, peut commettre des erreurs ou des inexactitudes qu'il aurait pu éviter autrement.

Nous faisons également l'hypothèse que le poste de pseudo-pilote induit bien souvent une forte charge cognitive. En effet, il implique de simuler plusieurs avions virtuels lors d'une même séance (parfois plus de 30 avions différents dans une même zone !). Rappelons que cela consiste à accomplir, pour chaque avion simulé, plusieurs tâches : répondre au contrôleur en fonction des paramètres de vol, déplacer l'avion selon les instructions données, etc. Or, comme nous le mentionnons plus haut, le nombre d'avions à gérer par un contrôleur joue sur sa charge cognitive ; nous postulons que la même conclusion s'applique pour le pseudo-pilote. Par contre le pseudo-pilote est un locuteur

---

<sup>37</sup> Selon sa charge cognitive, l'interlocuteur peut également rencontrer des problèmes dans le traitement de ce qu'il entend. Cependant, dans le cadre de cette thèse, nous nous plaçons essentiellement dans la perspective de la production (dans laquelle apparaissent les disfluences), et donc du seul locuteur.



maîtrisant la phraséologie ; de ce fait, conformément à notre hypothèse énoncée ci-dessus, nous postulons que sa charge cognitive doit être peu affectée par l'utilisation de la phraséologie.

#### 3.5.2. Caractéristiques linguistiques

Il est important de contrer une éventuelle objection qui pourrait nous être opposée. On peut, se demander si on a bien affaire ici à de l'oral spontané. Nous tenons à argumenter cette affirmation. En effet, le cadre très restrictif fixé par la phraséologie pourrait laisser à penser que tous les énoncés produits sont déjà planifiés à l'avance. Ce n'est pas le cas : ni les contrôleurs, ni les pilotes ne savent à l'avance ce qui va arriver, et par conséquent ce qu'ils vont avoir à prononcer. La phraséologie définit uniquement le cadre général de production des énoncés ; ce qui est énoncé repose sur l'interaction **dynamique** entre un contrôleur et un pilote ou un pseudo-pilote donné, en fonction d'une situation de vol donnée. Cela est également valable pour la situation d'apprentissage : comme nous l'avons expliqué (section 3.3), les conditions pour les exercices d'évaluation sont aussi proches que possible du contexte réel ; de ce fait, les contrôleurs en apprentissage ne savent pas plus à l'avance ce qui va se passer que leurs collègues en contact avec le trafic « réel ». Ceci est le principe même du fonctionnement du dispositif de pseudo-pilotage, et son principal intérêt en tant qu'exercice d'entraînement et d'évaluation.

Une autre conséquence de la spécificité de l'interaction du contrôle aérien est l'impossibilité d'occuper simultanément le même canal. Ainsi, un phénomène arrivant couramment dans l'oral spontané et pouvant perturber la communication, est le « recouvrement de parole » (*overlapping* en anglais). Ce phénomène se produit lorsque deux locuteurs ou plus parlent simultanément.

La phraséologie prévient également, ou du moins limite, l'occurrence de certaines spécificités de l'oral spontané, que nous avons présentées dans le chapitre 1.

En termes lexico-syntaxiques, comme nous l'avons déjà noté, l'usage de la phraséologie implique un certain nombre de caractéristiques qui ne se retrouvent pas dans les productions linguistiques non-contrôlées.

La phraséologie pose également, plus encore que dans le langage naturel non contraint, le problème de la détermination d'un « mot » comme unité minimale. Cela est surtout dû à la profusion des expressions des valeurs numériques telles que les niveaux, les vitesses, ou les indicatifs. Ce problème est accru par la nécessité de prononcer, dans certaines situations, chaque composant d'une valeur.

Soit, par exemple, le nombre « 122 ». Dans le contexte du contrôle aérien, il peut se prononcer de deux manières différentes. La première est « cent vingt deux ». Dans ce cas de figure, conformément à l'usage en français, il s'agit bien d'un seul mot (composé de trois éléments lexicaux). Il y a par contre problème avec le second mode de prononciation.

Il est employé lorsque la nécessité d'une prononciation claire s'impose. Dans ce cas, notre exemple « 122 » sera prononcé « unité deux deux ». Dans cette deuxième modalité de prononciation deux interprétations sont possibles quant au nombre de mots prononcés : un seul, ou trois.

En termes pragmatique, la principale différence avec le dialogue oral homme/homme ou homme/machine est le fait que les connaissances relatives à la tâche sont partagées par les interlocuteurs, avant même le début du dialogue. Comme indiqué dans le chapitre 1 section 2.3.3, ce partage des connaissances (*grounding*) joue un rôle important. Dans le cadre du contrôle aérien, les dialogues sont conçus de manière à limiter autant que possible la part des messages consacrés à l'établissement du *grounding*. Ce processus peut occuper une grande partie d'un dialogue dans d'autres types de tâches. Les processus de mise en commun des connaissances sont assurés au début de la conversation, avec l'identification du contrôleur et du pilote. Beaucoup des autres informations nécessaires sont données par l'instrumentation électronique : radar et instruments de vol. Il y a malgré tout des mises à jour de la base commune des connaissances qui sont doivent passer par la modalité orale. Mais leur importance est limitée : elles concernent uniquement les paramètres de vol ou sont liées à la situation géographique de l'aéronef considéré.

Un dernier point à noter est le fait que le débit de parole employé par les contrôleurs aériens est en général plus élevé que lors d'une production orale spontanée courante : environ un peu plus de 150 mots/minute en moyenne<sup>38</sup>. Nous ne disposons malheureusement pas d'études antérieures sur ce sujet. Nous n'avons que des hypothèses sur l'interprétation de ce taux de débit de parole. Nous pensons notamment au fait que la tâche de contrôle et de régulation doit être effectuée aussi rapidement que possible, comme nous l'avons décrit plus haut (3.1.1) ; d'autre part, on peut aussi attribuer cette rapidité à la connaissance partagée de la phraséologie acquise par les élèves contrôleurs que nous venons de mentionner : les dialoguants savent que leur interlocuteur a les mêmes connaissances qu'eux, et n'ont pas à parler lentement pour se faire comprendre. Le système de redondance introduit par le collationnement offre un « filet » potentiel en cas de mauvaise compréhension.

Quoi qu'il en soit, cette autre spécificité du corpus est intéressante. Nous voyons ainsi que la phraséologie et la tâche influencent, non seulement la structure lexico-syntaxique des énoncés, mais également la réalisation phonétique elle-même.

---

<sup>38</sup> Nous avons obtenu ce résultat en calculant le rapport entre le nombre de mots des corpus et la durée des enregistrements (hors périodes de silences, cf. section 4 pour plus de précisions).

#### 4. Corpus d'étude

Nous présentons dans cette section les deux corpus que nous avons utilisés dans le cadre de cette thèse. Ces corpus sont constitués de dialogues entre contrôleurs aériens et pilotes, mais ils diffèrent sur plusieurs points : conditions d'acquisition, caractéristiques techniques et enfin utilisation que nous en avons faite.

Le premier corpus, que nous appellerons corpus A, est celui que nous avons le plus utilisé pour les études de cette thèse. Nous l'avons-nous-même transcrit. Le deuxième corpus (que nous appellerons désormais corpus B) avait déjà été entièrement transcrit par un organisme tiers. Comme nous le verrons dans le prochain chapitre, ce détail a son importance. Il explique notamment que nous avons utilisé le corpus B principalement pour des études statistiques, et en complément de validation de certaines hypothèses. Toutes ces différences sont explicitées dans les sous-sections suivantes.

Ces deux corpus sont des recueils d'enregistrements de dialogues oraux spontanés ; mais notre travail a été réalisé sur les transcriptions de ces enregistrements. Ces transcriptions sont complétées par des annotations utilisées pour signaler des phénomènes autres que lexico-syntaxique, notamment les disfluences.

En conséquence, nous décrivons successivement les deux composantes: orale et écrite, de chacun de ces deux corpus.

##### 4.1. Composante orale

###### 4.1.1. Contexte d'acquisition

###### 4.1.1.1. Corpus A

Ce corpus est constitué d'enregistrements de dialogues oraux spontanés acquis, lors de sessions d'apprentissage, dans le cadre du projet VOICE<sup>39</sup>. Ce projet se situe dans la continuation d'autres projets menés au CENA sur l'interaction vocale. Avec le perfectionnement des techniques de traitement automatique de la parole, il a été envisagé d'utiliser la modalité vocale dans les applications liées à l'aéronautique, notamment des environnements de simulation du contrôle de trafic aérien. Deux principaux projets ont été consacrés à ces objectifs : *PAROLE*<sup>40</sup> (maquette d'une interface vocale) et *VOCALISES*<sup>41</sup> (programme d'industrialisation de la maquette). Malgré les applications potentielles, ces projets ont été abandonnés dans le milieu des années 90 en raison des faibles performances des techniques de reconnaissance vocale (cf. Marque *et al.* (1995, p. 14) et Maugis (1995, p. 20-21).

---

<sup>39</sup> *VOcal Interaction in Control Environment*. Initialement nommé VICTOR (Truillet & Vigouroux (2001)).

<sup>40</sup> *Pilotage Automatique par Reconnaissance des Ordres des contrôLeurs à l'Entraînement*

<sup>41</sup> *Voice Comprehension for Advanced Learning and Impress of Syntax through Extended Simulators*.

Suite aux récents développements matériels et logiciels des traitements du signal audio, l'idée a été relancée, avec le projet VOICE. L'objectif de ce projet est de récolter des corpus relatifs au contrôle du trafic aérien, de formaliser sous la forme de modèles de langage (comme par exemple dans McTait *et al.* (2004) et Dourmap & Truillet (2003)) la phraséologie utilisée (Maugis (1995)), et d'évaluer les résultats obtenus par une nouvelle maquette d'interface vocale de pseudo-pilotage. Les résultats obtenus sont encourageants (Truillet *et al.* (2005, p. 72)).

L'objectif final de ces études est d'utiliser toutes ces données afin de concevoir un environnement d'apprentissage et d'entraînement dans lequel les pseudo-pilotes seraient assistés par des agents dialoguants, dans le cadre des perspectives que nous avons présentée en section 3.5.1. Toute aide leur permettant de se décharger, même partiellement, d'une charge de travail naturellement trop forte est la bienvenue !

Les enregistrements ont été réalisés en juillet 2001, à l'ENAC de Toulouse.

#### *4.1.1.2. Corpus B*

Ce corpus appartient au DSNA<sup>42</sup>. Il a initialement été recueilli dans le cadre du projet VOCALISE. Le but était d'étudier les communications radio-téléphoniques relatives au contrôle du trafic aérien, et plus particulièrement de disposer de points de comparaison avec l'utilisation de la technologie Data Link.

Contrairement au corpus A, la situation d'enregistrement est « réelle » : il s'agit de communications concernant des vols réguliers de l'espace aéronautique. Comme dans le corpus A, il s'agit de communications réalisées dans des centres « en route » et non « en approche ».

Pour plus de détails, nous renvoyons à Graglia *et al.* (2005) ainsi qu'à la présentation faite sur le site web consacré : <http://www.cena.fr/divisions/ICS/projets/vocalise/>.

#### *4.1.2. Caractéristiques techniques*

##### *4.1.2.1. Corpus A*

16 contrôleurs en formation, de langue maternelle française, ont été enregistrés. Aucun ne souffre de troubles du langage. Les âges sont compris entre 20 et 30 ans.

Les enregistrements ont été réalisés avec un Digital Audio Tape (DAT). Le signal de parole des contrôleurs a été échantillonné à 16 kHz (16 bits). Etant donné les conditions difficiles d'enregistrement, la qualité acoustique souffre de problèmes dus à la saturation ou aux bruits tels que les interférences. Cependant, les dialogues restent intelligibles.

---

<sup>42</sup> Direction des Services de la Navigation Aérienne (Air Navigation Services Department).

Nous présentons les principales caractéristiques du corpus A dans le Tableau 2 ci-dessous.

**Tableau 2** : principales caractéristiques du corpus A

Durée moyenne d'un énoncé	Durée totale des dialogues	Nombre d'énoncés	Nombre de mots	Pourcentage d'énoncés en anglais	Pourcentage d'énoncés en français	Nombre de contrôleurs en formation
5,42s	36h50mns	11 427	76 306	48%	52%	16 (répartis en 2 groupes)

Pour comprendre le tableau, il est nécessaire de faire deux clarifications terminologiques. Nous appelons « énoncé » toute production orale d'un locuteur donné, dont les bornes sont définies soit par l'intervention d'un autre locuteur, soit par une coupure silencieuse de plus de 3 secondes<sup>43</sup>.

D'autre part, la définition de la notion de « mot », déjà difficile dans le cadre d'un langage non contraint, est encore plus difficile dans le cadre de la phraséologie. En effet, comme nous l'avons vu, celle-ci emploie très fréquemment des sigles et combinaisons de chiffres. De ce fait nous avons été contraints d'adopter la définition la plus large : on appellera « mot » toute séquence de caractères alphanumériques, dont les bornes sont fixées par une espace ou un signe de ponctuation. Ainsi, pour reprendre l'exemple de la section 3.5.2 ci-dessus, « 122 » sera considéré comme un seul mot, mais « 1 2 2 » (les espaces représentant le fait que le nombre est prononcé « unité deux deux ») comptera comme trois mots.

Le nombre donné dans le tableau doit donc être considéré en gardant en mémoire le fait qu'il recoupe aussi les sigles et éléments de sigles. Deux langues sont utilisées : le français (majoritaire) et l'anglais.

##### 4.1.2.2. Corpus B

Les données résultent de l'enregistrement de plus de 60 heures de dialogues de contrôleurs aériens et de pilotes. Le dispositif d'enregistrement est analogique ; de ce fait, la qualité sonore est malheureusement assez dégradée, notamment par la présence de nombreux bruits parasites, ainsi que par des effets de saturation du signal. A fins d'étude, les données analogiques ont été converties en fichiers .wav (sans compression), quantifiés à 8 bits avec un taux d'échantillonnage de 8 Khz. Le corpus B présente ainsi des caractéristiques d'acquisition et de compression moins performantes que pour le corpus A.

---

<sup>43</sup> Cette définition est quasi-identique à celle donnée dans Coullon *et al.* (2000).

Sur les 60 heures d'enregistrements récoltées, 13 heures de dialogues sont utilisables, après avoir enlevé les différentes zones de silence<sup>44</sup>, fréquentes dans ce type de communications.

Les caractéristiques de ce corpus sont présentées dans le Tableau 3 ci-dessous.

**Tableau 3** : principales caractéristiques du corpus B

Durée moyenne d'un énoncé	Durée totale des dialogues	Nombre d'énoncés	Nombre de mots	Pourcentage d'énoncés en anglais	Pourcentage d'énoncés en français
3,95s	13 h	12 546	120 000	57%	43%

Par rapport au corpus A, on observe que le nombre moyen de mots par énoncé est plus important pour le corpus B. La raison est que plus d'ordres complexes sont donnés dans la situation « réelle ».

## **4.2. Composante écrite**

### *4.2.1. Transcription et annotation*

#### *4.2.1.1. Corpus A*

En règle générale, toute méthode de transcription et d'annotation peut adopter un ou plusieurs des cinq niveaux suivants :

1. Orthographique : mettre à l'écrit ce qui est dit, ainsi que, si possible, les sons de l'environnement immédiat (par exemple des bruits d'arrière-plan). Ce niveau peut être augmenté par des balises permettant de représenter les phénomènes prosodiques et extra-linguistiques, disfluents ou non : toux, rires, pauses, hésitations, etc.;
2. Phonétique : transcrire ce qui a été dit en utilisant l'alphabet phonétique international (l'I.P.A. : *International Phonetic Alphabet*<sup>45</sup>). Ce niveau est plus particulièrement utile pour l'apprentissage de modèles acoustiques pour des systèmes de reconnaissance automatique, ou pour rendre compte des différentes prononciations d'un mot (par exemple dans le cas d'une prononciation par un locuteur étranger) ;
3. Grammatical : assigner des catégories grammaticales aux mots d'un énoncé. Par exemple, des analystes peuvent procéder à la *lemmatisation* des mots : tout mot infléchi (marque de genre, de nombre, conjugaison, etc.) est ainsi réduit à une forme canonique, appelée *lemme*. La lemmatisation n'est qu'une technique parmi d'autres à ce niveau ;

<sup>44</sup> Périodes intermédiaires entre deux dialogues pendant lesquelles l'enregistrement se poursuit malgré tout.

<sup>45</sup> Sur Internet, cf. <http://www.arts.gla.ac.uk/ipa/ipa.html> (page d'accueil de l'*International Phonetic Association*) et <http://www.arts.gla.ac.uk/ipa/ipachart.html> pour la table des différents phonèmes.

4. Sémantique : ce niveau peut être traité de différentes manières. Ainsi, on peut chercher à annoter des mots et/ou des énoncés en fonction de leur signification. Une autre possibilité offerte à l'annotateur est de centrer son intérêt sur les actes de langage (au sens de Austin (1962) et ses successeurs : J. Searle, D. Vanderveken) exprimés dans les énoncés à traiter. Dans le cas d'un corpus contenant des dialogues, tel que ceux que nous considérons ici, on peut également s'intéresser aux actes de dialogue (voir par exemple l'utilisation faite par Bunt (1996)). Ce dernier type de corpus peut également être annoté à un cinquième niveau, appelé dialogique ;
5. Dialogique : ce niveau concerne la structuration des énoncés produits par les participants d'un dialogue. Les méthodologies d'annotation s'appliquant à ce niveau sont en général inspirées des travaux visant à modéliser le dialogue ainsi que la combinaison de ses composants, tels que ceux que nous avons présentés dans le chapitre 1.

Comme nous l'avons décrit ci-dessus, il existe un ensemble de méthodes et de niveaux différents d'annotation. Le choix de l'une ou l'autre (ou de plusieurs) varie selon les buts de l'analyse linguistique du corpus. Dans le cadre du présent travail, une transcription et annotation des enregistrements audio ont été réalisés, par nous-mêmes, en suivant des spécifications définies dans (Coullon *et al.* 2000) et (Coullon *et al.* 2001). Les auteurs font une distinction entre les niveaux de transcription orthographique et d'annotation. L'annotation correspond à une interprétation (aux niveaux sémantique, dialogique, etc.) de la chaîne orthographique.

Le but des spécifications de la transcription est de déterminer les éléments à transcrire, d'obtenir l'homogénéité des transcriptions dans le cas où plusieurs transcripteurs réalisent cette tâche sur le corpus. Les spécifications consistent essentiellement en des règles à suivre pour transcrire les termes techniques tels que les indicatifs, les vitesses, etc. Elles donnent également des instructions de transcription des phénomènes comme les hésitations ou les pauses. La règle générale est :

**« Le principe de la transcription des énoncés est de respecter au maximum ce qui a été réellement dit à la fréquence, aussi bien sur le plan de l'énonciation (forme verbale) que sur le contenu. Aucune interprétation n'est effectuée à ce niveau. » (Coullon *et al.* ((2000, p.10))**

Nous avons complété ces spécifications de manière à prendre en compte les phénomènes prosodiques porteurs selon nous d'informations supplémentaires. Par exemple, nous avons ainsi relevé à plusieurs reprises un « appui » effectué par le locuteur sur des lexèmes tels que les indicatifs par exemple. Cependant, ce type de notation étant subjective, nous n'avons pas tenu compte dans les recherches que nous avons effectuées sur le corpus.

Le logiciel *Transcriber*<sup>46</sup> (version 1.4.2) a été utilisé pour les transcriptions. Ce logiciel est spécifiquement conçu pour permettre et faciliter autant que possible la transcription d'enregistrements audio de toute nature. En effet, il offre dans la même interface les fonctionnalités de transcription et d'écoute de l'enregistrement, avec possibilité de retour en arrière, de rejouer autant de fois que nécessaire un même passage, etc. L'interface permet également de rajouter dans la transcription un certain nombre de balises prédéfinies correspondant aux phénomènes pouvant advenir à l'oral spontané : bruits divers, onomatopées, etc. ; le cas échéant l'utilisateur peut aussi définir et rajouter des balises personnalisées. Nous avons utilisé cette fonctionnalité pour baliser les disfluences que nous avons relevées.

D'un point de vue fonctionnel, ce logiciel permet également de créer facilement des profils correspondant aux différents locuteurs. Il est ainsi possible de faire correspondre à chaque production linguistique l'identité du locuteur. Pour cela, le transcripateur crée un ou plusieurs identifiants, en fonction du nombre de locuteurs du corpus à transcrire. Comme nous ne connaissons pas l'identité des différents contrôleurs et pseudo-pilotes ayant participé aux enregistrements du corpus A, nous avons adopté une nomenclature pour nommer les profils. Le but de cette nomenclature est de résumer dans l'identifiant d'un locuteur donné les caractéristiques distinctives de celui-ci :

- Sa fonction (pilote ou contrôleur) ;
- Son sexe (homme ou femme) ;
- La langue parlée (anglais ou français). Dans le cas où un locuteur donné utilise les deux langues, deux profils différents lui seront affectés en fonction de la langue employée ;
- Eventuellement, un code numérique unique pour un locuteur donné, lorsque celui-ci est clairement identifiable.

Les codes correspondant à ces différentes caractéristiques sont combinés à l'aide du caractère « souligné » (« \_ »). L'application de cette nomenclature donne ainsi des noms de profils tels que « pil\_h\_ang\_2 » pour désigner un pilote masculin parlant en anglais, ou « ctrl\_f\_fr\_inde » pour un contrôleur de sexe féminin, parlant français, et n'ayant pas pu être identifié.

Le logiciel *Transcriber* offre aussi la possibilité de représenter de manière assez fine la hiérarchisation des dialogues, au sens que nous avons présenté au début de la section 4.2.1.1. Avec *Transcriber*, le transcripateur peut gérer les trois niveaux hiérarchiques suivants, classés par ordre croissant d'imbrication :

---

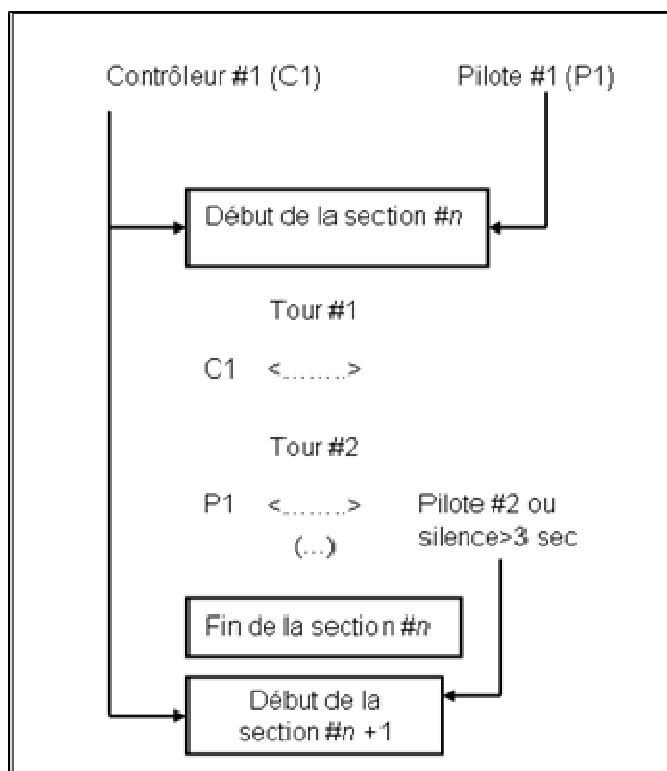
<sup>46</sup> Téléchargeable à l'URL : [trans.sourceforge.net/](http://trans.sourceforge.net/), et présenté pour la première fois dans Barras *et al.* (2000).



- Les énoncés : même définition que l'équivalent écrit, c'est-à-dire une unité du discours composé de différents mots ordonnés, et généralement marqués à l'écrit par une ponctuation, et à l'oral par des marqueurs lexico-syntaxiques, et prosodiques (pause longue, variation de la prosodie, etc.). Un locuteur peut enchaîner plusieurs énoncés à la suite ;
- Les tours de parole : ils correspondent à l'ensemble des énoncés produit par un locuteur donné jusqu'à l'alternance avec un autre locuteur. Chaque tour de parole peut contenir plusieurs énoncés, et est associé à un locuteur au moyen de l'identifiant donné par Transcriber ;
- Les sections : dans la structuration du dialogue, il s'agit du niveau supérieur à celui des énoncés et des tours de parole, qu'il englobe. Une section est un ensemble de tours de parole échangés entre un contrôleur donné et un pilote donné, au sujet d'un thème donné (par exemple, séries d'instruction dans l'approche d'un aérodrome, etc.). Du fait de la définition que nous avons donnée, les bornes d'une section seront constituées par au moins un des deux éléments suivants :
  - Changement d'un ou des deux locuteurs (pilote/contrôleur) : par exemple, le contrôleur s'adresse à un nouveau pilote entré dans le secteur qu'il gère, ou bien un pilote entre dans un nouveau secteur, géré par un autre contrôleur ;
  - Changement du thème de dialogue : un exemple typique est lorsqu'un pilote a reçu et collationné les instructions données par le contrôleur, et qu'il le rappelle pour recevoir de nouvelles instructions suite à un changement de sa situation (par exemple, changement de cap).

Nous avons défini trois types de sections différentes, correspondant à trois situations différentes d'enregistrement du corpus A. Il s'agit : des dialogues de contrôle aérien, des dialogues dépourvus d'intérêt (par exemple, calibrage du microphone, tests d'audibilité, etc.) mais que nous avons malgré tout transcrits, et enfin des périodes de silence (généralement des périodes intermédiaires entre les deux autres types de section). Chacun des enregistrements du corpus A appartient à l'une et une seule de ces catégories.

La Figure 9 ci-dessous représente la gestion des tours de parole et sections permise par Transcriber, telle que nous l'avons utilisée pour transcrire les dialogues de contrôle aérien.



**Figure 9 :** Hiérarchisation des sections et des tours de parole

Signalons également que Transcriber permet l'alignement des transcriptions sur le signal audio. Autrement dit, il est possible au transcripteur de faire correspondre à une unité donnée de la transcription (mot, balise, énoncé, tour de parole) l'instant de début et l'instant de fin de celle-ci, avec une précision de l'ordre du centième de secondes. Il est ainsi aisé de parcourir les transcriptions au moyen de paramètres temporels : intervalle de tel ou tel phénomène, position dans le signal, etc.

Enfin, les transcriptions sont enregistrées dans un format de fichier compatible XML<sup>47</sup>.

#### 4.2.1.2. Corpus B

A partir des échantillons audio, des premières transcriptions ont été effectuées par des pilotes. Ils étaient aidés pour cette tâche par la visualisation sur des écrans radars des indicatifs des vols impliqués. Ce dispositif était particulièrement utile compte tenu de la mauvaise qualité de l'enregistrement : en raison de celle-ci, certains mots sont difficilement intelligibles. L'écran radar a alors aidé à désambigüiser voire comprendre certaines occurrences.

Comme pour le corpus A, la transcription a été assurée grâce au logiciel Transcriber, selon les principes décrits précédemment.

<sup>47</sup> *eXtended Markup Language*

En ce qui concerne les conventions de transcription, elles diffèrent légèrement de celles utilisées pour le corpus A, sur les phénomènes à annoter et la manière de le faire. Cependant, là aussi le but était de relever différents phénomènes extra linguistiques.

### 5. Conclusion

Ce chapitre a été consacré à la présentation du contexte particulier des dialogues oraux spontanés que nous étudions : le contrôle aérien. Nous avons montré que la tâche de contrôle aérien présente des caractéristiques très spécifiques.

D'un point de vue linguistique, la principale caractéristique à retenir est l'utilisation de la phraséologie. Celle-ci concerne aussi bien le lexique utilisé que la syntaxe employée. Elle a pour fonction de permettre la réalisation de la tâche de contrôle, en prenant en compte les impératifs de la tâche, parmi lesquels les plus importants sont la rapidité d'exécution et l'intelligibilité. Le lexique utilisé comporte peu de lexèmes différents, et est particulièrement contraint par la tâche.

Nous avons montré que la particularité de la tâche a également de l'influence au niveau cognitif, puisqu'elle induit une très forte charge cognitive, qui est, par contre, compensée par le relativement faible nombre de schémas à maîtriser. Cette dernière remarque est surtout valable pour l'utilisation de la phraséologie en situation « réelle » mais sans doute moins pour celle de l'apprentissage des contrôleurs aériens. Nous avons présenté ensuite les principales caractéristiques de la phraséologie, ainsi que la manière dont l'apprentissage est réalisé.

A la suite de ces présentations, nous avons décrit les corpus de contrôle aérien qui nous ont servi de base de travail. L'un fut enregistré dans le cadre de l'apprentissage de la phraséologie par les contrôleurs, alors que l'autre se situe dans un contexte de travail « réel ».

Nous disposons à présent de toutes les données nécessaires pour analyser ces corpus. C'est l'objectif du chapitre suivant.

# **PARTIE 2**

## **ANALYSE ET MODELISATION**



## Chapitre 3

# Analyse statistique et interprétative des phénomènes

### 1. Introduction

Le premier chapitre de cette thèse a présenté un diaporama général des principales caractéristiques des disfluences dans le dialogue oral spontané. Dans le deuxième chapitre, nous avons exposé les principales caractéristiques de la tâche de contrôle aérien, et plus particulièrement la communication entre les pilotes et les contrôleurs, aux niveaux pragmatique, linguistique, et cognitif. Nous avons ainsi montré qu'il s'agit d'un type de tâche très particulier, et dont la spécificité se fait ressentir à tous ces niveaux.

Nous pouvons alors nous demander si celles-ci peuvent avoir des répercussions sur les disfluences, que ce soit au niveau du mode de manifestation, ou bien de leur fréquence. Plusieurs raisons nous amènent à postuler l'hypothèse d'une différence également à un ou plusieurs de ces niveaux. D'abord, le fait que le contrôle aérien soit une tâche très critique ; par conséquent, nous subodorons que les participants à ce travail sont particulièrement attentifs à ce qu'ils disent, et qu'ils évitent autant que possible toute source d'erreur ou d'ambiguïté, comme peuvent être les disfluences. D'autre part, il faut considérer le fait que les disfluences sont également un phénomène linguistique. Or, on sait maintenant que la phraséologie et la tâche ont des influences indéniables sur les aspects linguistiques des communications (cf. chapitre 2). Il est ainsi logique de supposer que la phraséologie a également des conséquences sur les disfluences. Cette hypothèse est par ailleurs confirmée par différentes études sur la littérature consacrée à ce sujet (cf. chapitre 1). Elles montrent que le contexte d'énonciation, au sens large, conditionne l'apparition des disfluences. Il s'agit aussi bien du contexte lexico-syntaxique (cf. par exemple Henry *et al.* (2004)), que de paramètres extra-linguistiques : Branigan *et al.* (1999) montrent que des facteurs tels que le sexe, le rôle du locuteur dans la conversation, la vision ou pas de l'interlocuteur, etc. affectent le taux de disfluences, et leurs types. Il est par conséquent envisageable qu'une forte modification du contexte de communication entraîne une modification des manifestations de disfluences. Or, le dialogue de contrôle aérien, avec sa dimension de communication particulièrement dynamique (gestion d'aéronefs en temps réel), sa phraséologie, est assurément un contexte de communication différent de celui traité dans la littérature.

La suite de ce chapitre est consacrée à la vérification de cette hypothèse, et à l'étude détaillée des particularités observées. Pour cela, nous présentons dans un premier temps la méthodologie de recherche que nous avons utilisée pour l'analyse du corpus. Nous décrivons ensuite les résultats obtenus, et menons une comparaison avec d'autres études sur les disfluences, effectuées sur d'autres types de corpus. Enfin, nous discutons les différents éléments que nous avons dégagés jusque là, en les replaçant dans la problématique de cette thèse.

## 2. Méthodologie d'analyse des disfluences

Dans cette section, nous présentons la démarche d'analyse que nous avons entreprise pour étudier la nature et les occurrences des disfluences à partir de l'analyse du corpus A (corpus de dialogues de contrôleurs aériens en formation, décrit dans le chapitre 2, section 4.1.1.1). Dans un premier temps, nous décrivons la méthode que nous avons mise en place. Nous présentons ensuite la typologie que nous avons définie pour catégoriser les disfluences. Enfin, nous détaillons les marqueurs que nous avons utilisés pour notre analyse.

### 2.1. Méthode générale

Comme nous le signalons dans le chapitre 2, nous avons transcrit et annoté nous-mêmes le corpus A. Au fur et à mesure de la tâche de transcription et d'annotation, nous avons comparé nos observations avec celles de la littérature existante, afin d'identifier les phénomènes de disfluences, et d'en dresser une typologie, que nous présentons en 2.2. Nous avons ensuite appliqué à chaque type de ces phénomènes des marqueurs spécifiques sous la forme de balises. Notre objectif est de permettre, dans un troisième temps, divers traitements informatiques à partir de ces balises. Ces traitements consistent essentiellement en des recherches de marqueurs effectuées au moyen de scripts écrits dans le langage *Perl*<sup>48</sup> afin d'obtenir un recensement ciblé des phénomènes, et des statistiques.

### 2.2. Typologie et terminologie

Nous avons vu dans le chapitre 1, la variation, selon les auteurs, de la typologie et de la terminologie utilisées pour désigner les disfluences. Avant de commencer nos recherches, nous définissons notre propre catégorisation pour ces phénomènes, compte tenu des buts poursuivis.

Nous avons défini six différentes catégories de disfluences. À celles-ci s'ajoutent évidemment les stratégies d'auto-corrections, auxquelles nous consacrons une section à part entière étant donné la complexité du sujet. Mais cette distinction est fondée

---

<sup>48</sup> *Practical Extraction and Report Language*.

uniquement sur un but de présentation : nous considérons les corrections comme une septième catégorie de disfluences.

Nous donnons des exemples pour illustrer notre propos (en mettant la disfluence en gras).

- Hésitation : désigne uniquement l'interjection « euh ». Exemple :

maintenons niveau 1 0 0 Poitiers Amboise **euh** Lacan

- Répétitions : un mot (ou un groupe de mots) apparaît au moins deux fois à la suite. Nous n'avons pas pris en compte la répétition de disfluences. Exemple :

station **station** calling euh repeat your callsign

- Fragment de mot : un ou plusieurs phonèmes inidentifiables (par opposition aux amorces). Exemple (le fragment est entre crochets) :

due to [**ou**] due traffic euh descend level 9 0

Dans cet exemple, il est impossible de déterminer ce qu'à voulu dire le locuteur, ni même si le fragment considéré est un début de mot ;

- Amorce : l'arrêt de la production d'un mot avant la fin normale de celui-ci. Dans notre terminologie, une amorce correspond toujours à un fragment de mot que l'on peut identifier (souvent grâce à la connaissance de la phraséologie). Le lecteur peut se demander comment nous pouvons être assurés que ces phénomènes correspondent effectivement à des mots susceptibles d'être produits sur les chaînes syntagmatiques ou paradigmatiques et non des problèmes d'articulation. L'explication repose sur les caractéristiques du langage de communication entre pilotes, et du contexte de tâche. Du fait de la limitation du lexique et du nombre de schémas syntaxiques, il est aisé pour un analyste expert du domaine de déterminer si un mot est mal prononcé ou inachevé, et dans ce dernier cas de quel mot il s'agit. Cette connaissance du domaine permet ainsi de déterminer si ce mot apparaît dans l'énoncé où la faute est commise. Si cela est effectivement le cas, on considèrera qu'il s'agit d'une amorce, comme expliqué en section 2.2. Exemple (l'amorce est entre crochets) :

speed euh 200 **Kts** [**mak**] euh minimum

Le contexte et la phraséologie aident à comprendre que le contrôleur commence à prononcer « maximum ». Il se rend compte que cela ne convient pas et s'interrompt (« mak »). Enfin, il dit le mot correct : « minimum » ;

- Allongement : l'allongement d'une unité phonétique d'un mot, supérieur à 0,2 seconde (le seuil de 0,2 seconde est motivé par le travail de Candéa (2004, *op. cit.*, p. 23 *et sq.*)) peut être combiné aux hésitations. Ce phénomène entre également dans la catégorie des « pauses remplies » ;



- Pause longue : toute pause supérieure à 0,2 seconde et comprise à l'intérieur d'un tour de parole.

Les phénomènes appartenant aux quatre premières de ces catégories sont caractérisés par des configurations lexico-syntaxiques qui leur sont spécifiques. Ces configurations servent de *marqueurs* lorsqu'elles sont utilisées pour une recherche de phénomène. Nous les présentons en détails dans la section suivante.

### 2.3. Marqueurs recherchés

Nous avons distingué trois différents types de marqueurs: lexicaux, intonationnels, et autres phénomènes du discours oral spontané. Les deux dernières catégories résultent de la modalité orale du corpus.

#### 2.3.1. Marqueurs lexicaux

Nous faisons des distinctions parmi les marqueurs lexicaux. Elles sont motivées par nos observations préliminaires lors de la transcription du corpus<sup>49</sup> :

- Déictiques : mots faisant référence à un élément déjà mentionné auparavant dans le discours. Dans le contexte du contrôle aérien, la configuration la plus fréquente est la suivante : “c’est IND”, où IND est un indicatif; par exemple :

c'est A M L 753

On notera que cette utilisation des déictiques apparaît fréquemment dans d'autres contextes, notamment par les pilotes pour se présenter ;

- Excuse : par exemple, « désolé », « excusez moi », etc. ;
- Négation : tout mot utilisé afin d'exprimer la négation de quelque chose, le plus commun étant « non » ;
- Correction : le mot « correction ». Nous l'avons placé dans une catégorie spécifique, car son utilisation est explicitement demandée par la phraséologie pour effectuer la correction d'un énoncé ; selon celle-ci, la correction doit ensuite être suivie par l'élément corrigé.

#### 2.3.2. Marqueurs intonationnels

Nous employons le terme *intonationnel* pour désigner l'emphase placée sur un mot par le biais d'une variation de la prosodie. Ainsi, lorsqu'un locuteur corrige un élément erroné dans un indicatif, il arrive que l'élément corrigé soit prononcé avec un accent particulier. Par exemple, dans

---

<sup>49</sup> La mise en œuvre d'une méthode de transcription et d'annotation est une tâche itérative. Au fur et à mesure du travail sur le corpus, le transcripateur découvre de nouveaux phénomènes ou difficultés, qu'il doit prendre en compte. Cette prise en compte s'effectue *a posteriori* sur la transcription déjà accomplie jusque là.

l'élément en gras, qui corrige une valeur fausse précédemment donnée, a été accentué par le locuteur. Cette distinction peut sembler peu fiable, étant donné qu'elle fait appel à l'oreille humaine, et n'a pas été validée par des études acoustiques sur le signal lui-même, en prenant comme critères des paramètres tels que les variations de la fréquence fondamentale de la voix<sup>50</sup>, de l'énergie (mesurée en décibels), etc. Il nous paraît cependant approprié de le mentionner ici. En effet, nos observations sur les intonations employées, pour marquer une correction par exemple, ont souvent été confirmées *a posteriori* par le fait que le phénomène relevé est bien corrélé à une correction.

### 2.3.3. Marqueurs de discours spontané

La désignation générale « marqueurs du discours spontané » regroupe les diverses balises utilisées pour marquer les disfluences que nous avons définies dans la section 2.2 (à l'exception des répétitions). Ces balises ont été insérées dans la transcription en utilisant les fonctionnalités offertes par le logiciel Transcriber. Leur position dans la transcription correspond au moment de leur apparition dans le signal, la transcription étant alignée sur celui-ci.

À présent qu'ont été définis les disfluences et leurs marqueurs respectifs, nous pouvons étudier leur distribution dans le corpus.

## 3. Présentation générale des résultats (Corpus A)

Étant donné le nombre de catégories de disfluences et de marqueurs étudiés, nous procédons en plusieurs étapes pour la présentation des distributions correspondantes.

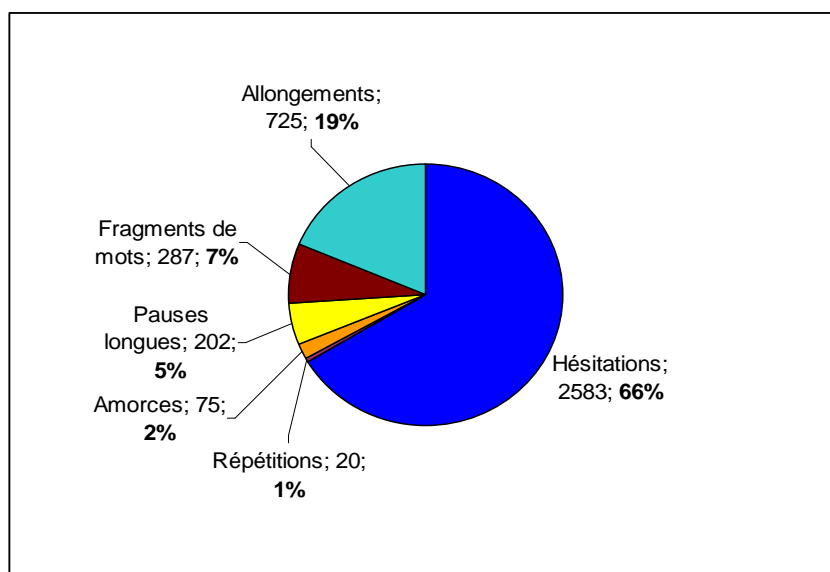
Dans une première sous-section, nous présentons les distributions relatives aux six principales catégories de disfluences, que nous avons définies en section 2.2. La deuxième section est consacrée entièrement à l'étude des phénomènes d'erreurs et de correction. Enfin, nous présentons dans la troisième section les distributions des marqueurs lexicaux-syntaxiques.

### 3.1. Distribution des phénomènes de disfluences

Nous donnons la distribution générale de l'ensemble des disfluences que nous venons de décrire. Nous la présentons en nombre total d'occurrences, et en pourcentage par rapport au nombre de mots du corpus d'étude. Ensuite, nous revenons en détails sur chacun de ces phénomènes, en les comparant, le cas échéant, avec les autres études similaires recensées dans le chapitre 1. Puisque chaque étude ne couvre pas l'ensemble des

<sup>50</sup> Désignée par « *F0* » en phonétique : la vibration émise par les cordes vocales avant passage dans les différents résonateurs que sont, entre autres, les cavités buccales et nasales.

disfluences, nous présenterons seulement celles qui concernent un phénomène donné, ou dont la catégorisation est proche de la nôtre, ou interprétable de la même manière<sup>51</sup>.



**Figure 10** : Distribution des disfluences dans le corpus A

#### 3.1.1. Répétitions

On voit dans le Tableau 1 que notre corpus comprend beaucoup moins de répétitions que les autres corpus. Notre hypothèse explicative est que, dans un contexte de contrôle aérien, les locuteurs (contrôleurs et pilotes) doivent être particulièrement vigilants afin d'éviter les ambiguïtés ou problèmes qui pourraient nuire à la compréhension de l'énoncé. Ce surcroît d'attention (par rapport à d'autres types de dialogues) permet d'éviter l'occurrence de répétitions disfluentes qui entraîneraient justement des ambiguïtés. De même, le temps nécessaire pour produire un énoncé n'est pas extensible : le locuteur ne peut pas perdre trop de temps en hésitation ou autres pauses (remplies ou silencieuses, telles que nous les avons définies dans le chapitre 1). Comme le montre la suite de ce chapitre, cette hypothèse est confirmée par le fait que, pour chacune des autres catégories de disfluences que nous avons définies, il y a systématiquement moins d'occurrences dans notre corpus (proportionnellement à la taille du corpus de comparaison).

Une autre particularité des répétitions disfluentes de notre corpus concerne leur structure. Nous avons vu dans le chapitre 1, section 3.2.4.4, que les répétitions disfluentes sont composées de deux blocs, appelés *répétable* (l'unité faisant l'objet de la répétition) et *répété* (la ou les répétitions du répétable. On sait que dans des dialogues non contraints (homme-homme ou homme-machine), le nombre de répétés peut varier de 1 (le répétable est répété 1 fois) à 6 (le répétable est répété 6 fois), voir par exemple Henry (2002, p. 475).

<sup>51</sup> Lorsque cela est possible, nous indiquons dans les tableaux le nombre d'occurrences et le pourcentage. Nous mettons en gras cette dernière mesure, afin d'en faciliter la lecture.

Or, notre corpus ne comporte aucune répétition dont le nombre de répétés soit supérieur à 1.

De même, considérons à présent le nombre d'éléments répétés. Henry *et al.* (2004, p. 79) rapportent que 17,7% des répétitions concernent plus d'un répétable (autrement dit, c'est plus d'un mot unique qui est répété). Dans notre corpus, nous n'avons relevé que 3 répétitions comportant 2 répétables, et aucune avec un nombre supérieur de répétables.

**Tableau 4** : Comparaisons pour les répétitions

Nom de l'étude	Notre corpus	Candéa (2004)	Shriberg (1994)	Kurdi (2003)
Nombre et/ou pourcentage de répétitions (par rapport au nombre total de mots)	20 / <b>0,03%</b>	110	141 / <b>1/%</b>	256 / <b>0,49%</b>

### 3.1.2. Hésitations

Le Tableau 5 confirme notre hypothèse, à savoir qu'il y a également bien moins d'hésitations dans notre corpus que dans ceux des autres travaux référencés. Ainsi, il y a 544 occurrences dans Candéa (2004), mais le corpus utilisé dans cette étude ne dure que 70 minutes, contre 35 heures pour le nôtre. De ce fait, il y a proportionnellement moins d'occurrences dans le corpus de Candéa (2004). On note cependant que la différence est globalement moindre que celle que nous avons observée pour les répétitions (cf. section 3.1.1 ci-dessus).

**Tableau 5** : Comparaisons pour les hésitations

Nom de l'étude	Notre corpus	Candéa (2004)	Kurdi (2003)
Nombre et/ou pourcentage d'hésitations (par rapport au nombre total de mots)	2583 / <b>3.38%</b>	544	3512 / <b>6.75%</b>

### 3.1.3. Amorces et fragments de mots

Henry *et al.* (2003) est la seule étude dont la catégorisation est la plus proche de la nôtre en ce qui concerne les « amorces » et « fragments de mots ». Elle présente également des statistiques détaillées sur leur distribution. Comme les auteurs ne font pas la distinction entre les « amorces » et les « fragments de mots », nous additionnons les occurrences de ces deux types de phénomènes qui apparaissent dans notre corpus. Il en résulte un total de 362 occurrences, soit 0.47% du nombre total de mots. Henry *et al.* (2003) fait état d'un total de 6094 occurrences des « fragments de mots » pour environ un million de mots (soit approximativement 0.6%). La distribution dans notre corpus de cette double catégorie est

### 3. Présentation générale des résultats (Corpus A)

assez proche de celle observée dans Henry *et al.* (2003), contrairement à ce que nous avons constaté pour les catégories précédentes. Une explication à ce résultat pourrait être le fait que notre double catégorie ne correspond pas exactement à celle définie par Henry *et al.* (2003).

#### 3.1.4. Allongements

Comme on le voit dans le Tableau 6, la fréquence d'apparition des allongements est moindre dans notre corpus que dans les autres.

**Tableau 6** : Comparaisons pour les allongements

Nom de l'étude	Notre corpus	Candéa (2004)	Shriberg (1994)
Nombre et/ou pourcentage d'allongements (par rapport au nombre total de mots)	725 / <b>0.9%</b>	284	669 (y compris "euh") / <b>7.9%</b>

#### 3.1.5. Pauses longues

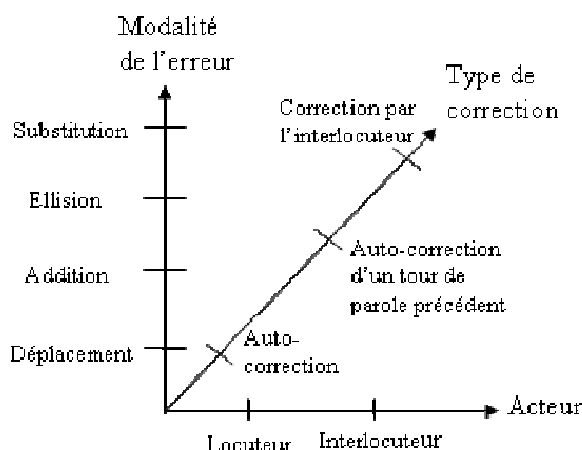
Le Tableau 7 montre que la spécificité de notre corpus est un peu moins prononcée que pour les autres catégories de disfluences. Cependant, là encore, on remarquera qu'il y a moins de « pauses longues » que dans d'autres corpus.

**Tableau 7** : Comparaisons pour les pauses longues

Nom de l'étude	Notre corpus	Candéa (2004)	Shriberg (1994)
Nombre et/ou pourcentage of pauses longues (par rapport au nombre total de mots)	827 / <b>1.08%</b>	1471	318 / <b>3.74%</b>

### 3.2. Distribution des phénomènes d'erreurs et de correction

Dans cette section, nous présentons la distribution des d'erreurs et de leur correction que nous avons observée, selon les mêmes méthodes de calcul que dans la section précédente. Nous les classons selon les trois niveaux différents d'interprétation de l'énoncé oral auxquels ils correspondent. Ces niveaux sont le rôle des participants au dialogue (locuteur ou interlocuteur) qui les produisent, les modes de manifestation de l'erreur, et le type de correction. Cette typologie correspond à la définition d'un espace de caractérisation des erreurs de performance orale. Les trois axes en sont schématisés dans la Figure 11 : ils vont d'une description macroscopique des stratégies palliatives à des unités phonétiques segmentales.



**Figure 11** : Espace de caractérisation d'une erreur de performance

### 3.2.1. Acteurs du dialogue et catégories de corrections

Nous faisons une distinction entre trois principales catégories de correction : l'auto-correction d'un élément de l'énoncé qui est corrigé, celle d'un énoncé précédent, ou la correction venant de l'interlocuteur. Les traits distinctifs de ces catégories sont basés sur la personne qui fait la correction (locuteur ou interlocuteur) et le tour de parole pendant lequel celle-ci se produit. L'analyse linguistique montre que les corrections se manifestent par des marqueurs spécifiques (entités lexicales et/ou éléments prosodiques), indépendamment les uns des autres, ou en même temps.

Voici des exemples de chacune de ces catégories, extraits de notre corpus (nous mettons l'élément corrigé en italique) :

- Auto-correction :

**Contrôleur** : KLM euh 2 1 5 montez niveau 1 9 0 contact ENAC 120  
contact ENAC euh 1 2 6 décimale 8 5

Le contrôleur demande au pilote d'aller au niveau 190, et de contacter l'ENAC sur la fréquence 126.85 : il effectue une correction sur la fréquence à utiliser, et le fait avant même d'avoir achevé sa phrase ;

- Auto-correction d'un énoncé précédent : voici un court dialogue entre un contrôleur et un pseudo-pilote :

**Contrôleur** : euh F K C correction maintenez niveau 1 7 0

**Pseudo-Pilote** : vers niveau 1 7 0 K C

**Contrôleur** : euh F K C correction maintenez niveau 1 9 0

Le contrôleur donne d'abord une position à laquelle le pseudo-pilote doit se rendre ; ce dernier confirme, mais, ensuite, le contrôleur corrige sa commande précédente, qui comportait des coordonnées erronées ;

- Correction par l'interlocuteur : un exemple de dialogue :

### 3. Présentation générale des résultats (Corpus A)

**Contrôleur** : euh TAT 289 M L (...) joignez Poitiers

**Pseudo-Pilote** : Lacan Amboise Poitiers c'est TAT M I

Dans cet exemple, le contrôleur produit une mauvaise énonciation de l'indicatif donné par son interlocuteur. En conséquence, celui-ci le corrige.

Pour la suite de cette étude, nous ne prenons en compte que les deux premières catégories. En effet, nous considérons que les corrections faites par un interlocuteur ne permettent pas de rendre compte directement de la manière dont le locuteur s'aperçoit par lui-même qu'il a commis une erreur, et dont il s'y prend pour y remédier. Or, c'est précisément ce que nous cherchons à étudier. De plus, elles ne posent aucun problème puisqu'elles apparaissent lorsque l'erreur a été notée par le locuteur aussi bien que par l'interlocuteur.

Le Tableau 8 ci-dessous présente la distribution des trois types de correction que nous venons de mentionner. Le pourcentage est donné par rapport au nombre de mots dans le corpus. On observe que le taux d'auto-correction est le plus important.

**Tableau 8** : Nombre et pourcentage des stratégies de correction

	Nombre	Pourcentage
Auto-Correction	146	2,29%
Auto-Correction d'un énoncé précédent	10	0,17%
Correction par l'interlocuteur	5	0,08%
Total	151	2,54%

A la catégorie des auto-corrections s'ajoute un type particulier d'amorce (définies plus haut en section 2.2). L'analyse a montré l'existence de deux types d'amorces. Elles diffèrent selon leur fonction dans l'énoncé. Le premier type n'a aucune fonction apparente. Exemple :

route Lacan [**amboi**] Amboise Balon Limoges

Ce type d'amorce peut être considéré comme une marque du « travail de formulation » pour reprendre le terme de Candea (2000). Il concerne 29 occurrences, c'est-à-dire 39% du nombre total d'amorces.

Au contraire, le second type correspond à une correction. Soit, par exemple, l'énoncé suivant :

[**mike**] Paris 124 decimal 05 Littoral M C

Ceci concerne 46 occurrences, c'est-à-dire 61% du nombre total d'amorces.

## 3.2.2. Manifestations de surface de l'erreur (niveau phonétique)

Ce que nous désignons par *manifestation de surface de l'erreur* recouvre les différentes façons dont un mot peut être mal prononcé. Pour en rendre compte, nous avons repris à notre compte la typologie classique utilisée dans les études psycholinguistiques, et présentée notamment dans Valdois (1994, p. 360-361), ainsi que Rossi *et al.* (1983, p.28-42). Les exemples donnés en illustration sous forme phonétique sont tous tirés de notre corpus ; la chaîne de phonèmes faisant l'objet d'une faute est mise en italique :

- Substitution : un phonème constituant le mot est substitué à un autre qui peut être présent ou non dans son contexte d'apparition. Exemple :

[novãbr(ə)nœfsãfaksfɔksromeo]

Ici, l'erreur résulte du fait que le locuteur, voulant produire la chaîne de phonèmes [fɔks], y insère le phonème [a] en lieu et place du [ɔ] ;

- Elision : un phonème constituant le mot est supprimé. Exemple :

[uisɛl(ə)lio l(ə)litoralzerovẽdœ]

[lio] correspond au début de [litoral], dont la production est interrompue car le locuteur se rend compte qu'il a omis le phonème [t] ;

- Addition : un (ou plusieurs) phonème(s) supplémentaire(s) est/sont ajouté(s) au mot produit. Il peut appartenir au contexte d'apparition. Exemple :

[gɔlfɔkszulutrafikœtuɛnœtɛn]

On observe deux productions du mot « ten ». La première est erronée en raison de l'addition du phonème [u]. Cela est rectifié par le locuteur qui produit ensuite la deuxième occurrence, correcte cette fois-ci ;

- Déplacement : la position de deux des phonèmes constituant le mot est intervertie. Exemple :

[tãgoynifɔrmremoromeo]

La chaîne de phonèmes en italique correspond à la production du début de l'indicatif [romeo]. Celle-ci est interrompue avant la fin, car le locuteur s'aperçoit qu'il a échangé la position de [o] et [e].

Ces manifestations de surface sont distribuées comme suit.



**Tableau 9** : Distribution des manifestations de surface (nombre et pourcentage)

	Addition	Omission	Substitution	Déplacement
Nombre d'occurrences	11	1	5	0
Pourcentage	64,71%	5,88%	29,41%	0%

### 3.2.3. Catégorisation des parties du discours affectées

En général, les analyses des erreurs dans la littérature du Traitement Automatique du Langage Naturel prennent comme critère de classification des composants touchés par les erreurs, les « parties du discours ». Autrement dit, les catégories morpho-syntaxiques : verbe, préposition, syntagme nominal, etc. en fonction de la granularité recherchée. Dans le cadre de notre étude nous ne choisirons pas ce critère de classification, qui nous aurait pourtant permis une comparaison plus précise avec les autres études. Il y a deux raisons à cela. Premièrement, la tâche de contrôle aérien présente certaines spécificités marquées, qui entraînent d'importantes modifications sur le langage naturel oral (cf. chapitre 2, sections 3.2 et 3.5.2). De ce fait, la répartition des catégories morpho-syntaxiques serait beaucoup moins pertinente à appliquer, d'autant plus qu'elle nécessiterait vraisemblablement des réajustements compte tenu des spécificités de notre corpus.

D'autre part, nous pensons qu'en termes explicatifs (quant à l'origine psycholinguistique des erreurs), une catégorisation morpho-syntaxique ne serait pas assez efficace. Nous formulons, en effet, l'hypothèse que certains des composants de la phraséologie, en particulier les identifiants, sont plus susceptibles que d'autres de faire l'objet d'erreurs, en raison de la difficulté de les énoncer oralement. En effet, les identifiants, composés d'une suite de lettres et de chiffres, ne correspondent pas aux mots du lexique courant : de ce fait, leur prononciation est plus difficile à la fois à planifier et à exécuter.

L'analyse linguistique du corpus révèle les composants des énoncés des contrôleurs susceptibles de faire l'objet d'une erreur. Nous avons distingué quatre sous-catégories, en fonction précisément de la nature de l'erreur qui provoque l'interruption de la production. Ces catégories sont définies selon la fonction occupée par le syntagme/mot qui fait l'objet de l'erreur. Cette approche, de ce fait, se rapproche de celle de Falzon (1982, p. 18-27) : dans les deux cas, la catégorisation se fait par rapport au rôle joué par l'unité dans l'exécution de la tâche. Par contre, la différence est que les catégories dressées par Falzon sont de granularité plus large, puisqu'elles concernent des énoncés entiers ou des segments

de taille importante<sup>52</sup>. Les unités que nous définissons ici sont de taille plus modeste. Elles correspondent aux unités affectées par une disfluente donnée. Comme nous le verrons, ce n'est jamais plus de 2 mots au maximum qui sont ainsi concernés.

La même remarque peut être faite en ce qui concerne la comparaison avec la décomposition très fine opérée par Volpe *et al.* (2003, p. 15-17). La hiérarchie qu'ils dressent comporte trois niveaux de granularité, et les niveaux les plus fins peuvent comporter plus d'une dizaine d'éléments différents. Cette précision dans la description est motivée, dans les deux études, par la volonté de rendre compte exhaustivement et finement de tous les éléments de sens intervenant dans la phraséologie. Or, notre objectif est différent : la catégorisation que nous entreprenons vise à identifier le contexte d'apparition des disfluences ; il est donc approprié qu'elle soit de granularité moins fine.

Les quatre catégories que nous avons définies sont les suivantes :

- Erreur sur un « mot » : nous appelons « mot » les données alpha-numériques (indicatifs, par exemple), et les commandes (« grimpez », « demande », etc.).

Exemple :

**Pilote** : climbing for level 1 7 0 and 2 0 0 Kts [mak] euh  
minimum D M C

- Erreur sur l'organisation de l'énoncé : un mot ou un groupe de mots n'occupant pas sa position correcte dans l'énoncé. Nous pourrions également employer le terme d'erreur « syntaxique », puisque ces erreurs concernent l'ordonnement des composants de l'énoncé, et leurs relations. Cependant, comme nous l'avons expliqué, la syntaxe de la phraséologie est très différente de celle du français « courant ». Pour éviter toute ambiguïté, nous préférons ne pas utiliser ce terme.

**Contrôleur** : [poi] Absie Poitiers Balon Reson Britair B X

Dans cet exemple, le contrôleur entame la prononciation du nom « Poitiers » ; or, ce nom de ville est inclus dans une liste des différentes étapes par lesquelles doit passer le vol auquel il s'adresse. Il se corrige donc en reprenant cette liste depuis son début.

- Erreur sur la langue utilisée : le locuteur remarque (ou bien on lui fait remarquer) qu'il n'a pas parlé dans la langue appropriée : français à la place de l'anglais ou vice versa.

**Contrôleur** : P I [vite] speed 2 1 0 Kts

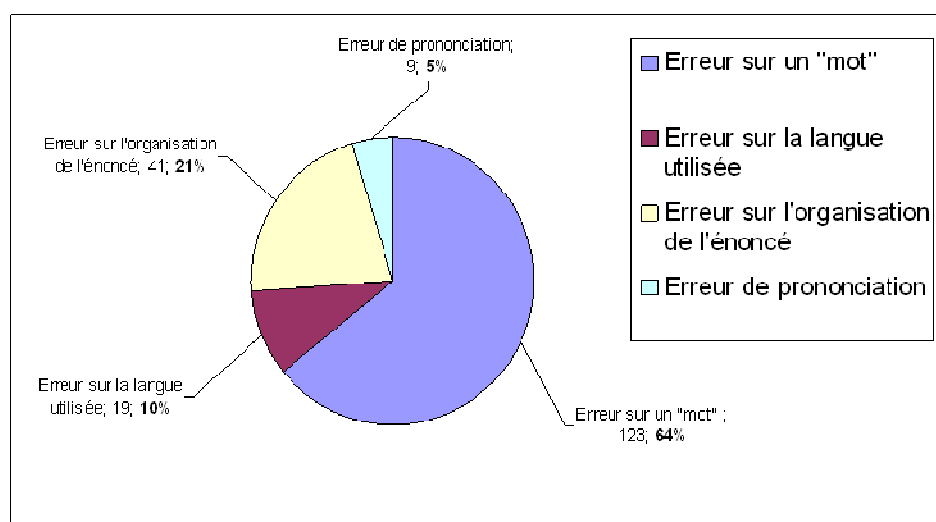
<sup>52</sup> Nous ne parlons pas de *syntagmes* car les segments considérés par Falzon ne correspondent pas à la définition linguistique du terme.

Ici, le locuteur commence à prononcer le mot « vitesse » ; or, il se rend compte qu'il doit parler en anglais, compte tenu du contexte (que nous n'avons pas reproduit ici). Il s'interrompt alors pour reformuler le mot en anglais.

- Erreur de prononciation : comme son nom l'indique, il s'agit d'un problème dans la prononciation d'un mot. Cette catégorie se définit principalement par opposition à celle que nous avons nommée « erreur sur un mot ». En effet, contrairement à cette dernière, elle ne désigne pas une erreur sur le choix du mot à employer, mais uniquement la production de celui-ci. Cette distinction nous paraît très importante quant à ce qu'elle sous-tend au niveau cognitif et psycholinguistique, comme nous le verrons en section 5.2 ci-dessous. Enfin, notons que l'unité concernée par ce type d'erreur ne pourra qu'être un mot.

**Pilote** : c'est le [lio] Littoral

La Figure 12 présente la distribution de ces catégories.



**Figure 12** : Distribution des corrections selon la fonction de la partie affectée

La plupart des erreurs concernent les références à des valeurs numériques : indicatif, vitesse, etc. Cela n'est guère surprenant car presque tous les énoncés contiennent au moins une référence à un indicatif, une vitesse, etc. Le même raisonnement peut s'appliquer aux erreurs portant sur les commandes. Cependant, il y a moins d'erreurs faites sur ces dernières que sur les références aux valeurs numériques. Cela est explicable par le fait que ces dernières, en particulier les indicatifs et les positions, sont des séquences assez complexes de combinaisons de nombres et de lettres, utilisées uniquement dans le contexte du contrôle du trafic aérien.

Par conséquent, le processus de production orale de ces valeurs nécessite certainement d'utiliser une charge cognitive de planification de diction et de mémorisation plus importante. Ainsi plus d'erreurs sont susceptibles de se produire durant ce processus. Nous

pensons aussi que ceci peut expliquer le nombre restreint d'erreurs portant sur les commandes (presque deux fois moins d'occurrences que pour les références à des valeurs) et d'organisation (plus de six fois moins d'occurrences que pour les références à des valeurs). Nous reviendrons plus bas sur ce point, en section 5.2.

Nous remarquons également que les erreurs ne concernant pas des éléments lexicaux (autrement dit les erreurs portant sur l'organisation de l'énoncé ou la langue utilisée : français ou anglais) sont largement minoritaires. Ce résultat peut être interprété à la lumière de ce que rapporte Reason (*op. cit.*, p. 156) ; selon lui, les erreurs d'ordre lexical sont plus sujettes que d'autres à échapper au module « d'édition » (cf. chapitre 1, section 3.1.2, pour une description de ce module).

Voyons à présent comment se distribuent les différents types d'erreurs affectant un mot, puisqu'il s'agit de la catégorie la plus affectée.

**Tableau 10** : Distribution des différents types d'erreurs touchant un mot (nombre et pourcentage)

	Erreurs concernant un mot entier	Erreurs pendant la production d'un mot, dues à une amorce	Erreurs pendant la production d'un mot, dues à une mauvaise articulation
Nombre d'occurrences	193	26	17
Pourcentage	80,08%	10,79%	9,13%

Le Tableau 10 reprend les deux principales catégories d'erreur que nous avons observées, à savoir les erreurs dues aux amorces et celles touchant l'ensemble d'un mot.

### 3.3. Fréquence et distribution des marqueurs

Dans un premier temps, nous présentons la distribution des différents marqueurs, définis plus haut en section 2.3, dans le corpus. Nous observons que la plupart apparaissent dans le contexte de corrections et auto-corrections. Cette observation va orienter le contenu des sous-sections qui suivent. Nous cherchons ensuite à déterminer plus précisément quelles sont les corrélations entre les marqueurs et les catégories de disfluences. Enfin, la dernière est consacrée à la présentation de la corrélation entre les marqueurs et les catégories de discours que nous avons présentées en section 3.2.3.

#### 3.3.1. Distribution globale des marqueurs

Les résultats sont donnés en nombre d'occurrences et en pourcentage (calculé en comparaison avec le nombre total de marqueurs) dans le Tableau 11.

**Tableau 11** : Nombre et pourcentage des marqueurs (toutes catégories confondues)

	Nombre	Pourcentage
Lexicaux	25	17,99%
Intonationnels	16	11,51%
Unités du discours oral spontané	98	70,50%
<b>Total</b>	<b>139</b>	<b>100,00%</b>

Quelques remarques avant de commenter ces résultats. Premièrement, ces marqueurs, spécialement ceux appartenant aux catégories intonationnelles et aux unités du discours oral spontané, apparaissent assez fréquemment dans le corpus, et la plupart du temps dans le seul contexte des corrections. En conséquence, les résultats doivent être considérés avec l'hypothèse que, lorsque les marqueurs se situent dans le contexte d'un phénomène de correction, ils lui sont directement liés, et peuvent ainsi être considérés comme son marqueur. On peut également prendre en compte le fait qu'il peut y avoir différents marqueurs pour une seule occurrence de correction ; mais cela n'affecte pas la distribution donnée dans le Tableau 11.

Le point le plus remarquable est le fait qu'il y a moins de marqueurs que d'occurrences de correction. Autrement dit, les corrections peuvent être identifiées même en l'absence de tout marqueur. Comme nous le montrerons dans le chapitre 4, cela est dû au fait que les corrections, même sans être signalées explicitement, présentent des patrons reconnaissables, une fois encore, à un expert du domaine de la tâche de contrôle aérien.

Penchons-nous à présent sur la distribution des différents types de marqueurs lexicaux dans le contexte des corrections (le pourcentage est calculé en comparaison avec le nombre total d'occurrences).

**Tableau 12** : Nombre et pourcentage des marqueurs lexicaux

	Nombre	Pourcentage
Déictiques	1	4,17%
Excuses	5	20,83%
Négations	4	16,67%
Corrections	14	58,33%
<b>Total</b>	<b>24</b>	<b>100%</b>

On voit que le marqueur lexical le plus fréquent est indubitablement le mot "correction", ce qui est évidemment attribuable au respect des règles de la phraséologie.

## 3.3.2. Corrélation entre marqueurs et types d'erreurs et de correction

**Tableau 13** : Distribution des différents marqueurs en fonction des catégories de correction (nombre d'occurrences et pourcentage)

	Auto-Correction	Auto-Correction d'un énoncé précédent	Amorce corrective	Correction par l'interlocuteur
Lexicaux	13 / <b>12,5%</b>	9 / <b>90%</b>	1 / <b>4,5%</b>	2 / <b>66,7%</b>
Intonationnels	13 / <b>12,5%</b>	1 / <b>10%</b>	1 / <b>4,5%</b>	1 / <b>33,3%</b>
Discours oral spontané	78 / <b>75%</b>	0 / <b>%</b>	20 / <b>90,9%</b>	0
<b>Total</b>	<b>104</b>	<b>10</b>	<b>22</b>	<b>3</b>

Le résultat le plus intéressant concerne la catégorie « auto-correction d'un énoncé précédent ». Le fait que presque tous les marqueurs de celle-ci soient lexicaux indique le respect de la phraséologie, qui stipule que la correction d'un énoncé précédent comportant une erreur doit être indiquée par le mot « correction ». Cela apparaît plus clairement lorsqu'on considère la catégorie « d'auto-correction » : tous les types de marqueurs peuvent apparaître dans son contexte, les plus fréquents étant ceux appartenant à la catégorie du discours oral spontané. Une explication pourrait être que, puisque la correction se produit durant l'énoncé courant, le locuteur réagit immédiatement, sans penser à respecter la phraséologie.

**Tableau 14** : Distribution des marqueurs lexicaux en fonction des catégories de correction (nombre d'occurrences et pourcentage)

	Auto-Correction	Auto-Correction d'un énoncé précédent	Amorce corrective	Correction par l'interlocuteur
Déictiques	0	0	2 / <b>38,5%</b>	0
Excuses	5 / <b>38,5%</b>	0	0	1 / <b>38,5%</b>
Négations	5 / <b>38,5%</b>	0	0	0
Corrections	3 / <b>23,1%</b>	9 / <b>100%</b>	0	0
<b>Total</b>	<b>13</b>	<b>9</b>	<b>2</b>	<b>1</b>

Deux phénomènes doivent être soulignés. D'une part, une fois encore c'est dans la catégorie des auto-corrections que les différents marqueurs lexicaux sont distribués de la manière la plus équilibrée. D'autre part, nous remarquons le phénomène opposé pour les « auto-corrections d'un énoncé précédent » : toutes les occurrences sont marquées par le mot « correction », ce qui peut être directement attribué au respect de la phraséologie, comme pour les résultats du Tableau 13. La distribution relative aux amorces et aux corrections concerne trop peu d'occurrences pour être interprétable correctement.

#### 4. Présentation générale des résultats (Corpus B)

##### 3.3.3. Corrélation entre marqueurs et catégories de discours

**Tableau 15** : Distribution des différents marqueurs en fonction des catégories de discours (nombre d'occurrences et pourcentage)

	Sur une valeur	Sur une commande	Sur l'organisation	Sur le langage utilisé
Lexicaux	20 / <b>22,5%</b>	3 / <b>8,8%</b>	0	2 / <b>40%</b>
Intonationnels	11 / <b>12,4%</b>	4 / <b>11,8%</b>	1 / <b>9,1%</b>	0
Discours oral spontané	58 / <b>64,2%</b>	27 / <b>79,4%</b>	10 / <b>90,9%</b>	3 / <b>60%</b>
<b>Total</b>	<b>89</b>	<b>34</b>	<b>11</b>	<b>5</b>

Le Tableau 15 montre que la proportion la plus importante des marqueurs est corrélée avec l'expression de valeur. Ce n'est pas une surprise : nous avons vu en section 3.2.3 que cette catégorie de discours est elle-même fortement corrélée avec les corrections.

#### 4. Présentation générale des résultats (Corpus B)

Cette section est consacrée à la présentation des distributions observées dans le corpus B. Dans un premier temps, nous présentons les phénomènes que nous avons pris en compte dans ce corpus. Dans un deuxième temps, nous présentons les résultats observés.

##### 4.1. Typologie des phénomènes recensés dans ce corpus

Comme nous l'indiquons dans le chapitre précédent, le corpus B avait déjà été annoté quand nous l'avons traité afin d'en examiner les disfluences. De ce fait, les phénomènes à prendre en compte avaient déjà été balisés en amont. Voici une liste exhaustive des différentes catégories prises en considération, chacune étant marquée par une balise spécifique :

- Les fautes de grammaire ou des ajouts incongrus effectivement prononcés ;
- Les mots coupés sur l'enregistrement (coupures dues à un problème technique) ;
- La présence de bruit ou de respiration ;
- Le fait qu'un mot ou groupe de mots ait commencé avant le début de l'enregistrement, ou en fin d'enregistrement ;
- Le fait qu'un ou plusieurs mots aient été prononcés mais qu'il a été impossible de les reconnaître ;
- Les hésitations (dans le même sens que celui de la terminologie que nous avons adoptée).

Nous n'avons pas pris en compte les trois premiers éléments car ils proviennent soit de problèmes de transcription, soit de « problèmes techniques » indépendants de la production orale elle-même. Ce n'est pas le cas des quatrième et cinquième éléments; nous les avons cependant ignorés, puisqu'ils ne correspondent pas à la disfluente telle que nous l'avons définie. Nous n'excluons cependant pas l'intérêt qu'ils pourraient présenter. Par contre, nous avons évidemment utilisé tout marqueur appartenant à la dernière catégorie. Enfin, nous avons par ailleurs recherché les mêmes marqueurs lexicaux que pour le corpus A.

#### **4.2. Résultats obtenus**

En préambule, il faut rappeler que le corpus contient approximativement 120 000 mots, et que le calcul a été fait après « nettoyage » du corpus des catégories 1 à 5 des conventions de transcriptions décrites dans le chapitre 2.

Nous présentons dans un premier temps les résultats relatifs aux disfluences. Nous nous intéressons ensuite à ceux qui concernent les phénomènes de correction. Dans chaque partie, nous commentons les principaux points pertinents.

##### *4.2.1. Fréquence des disfluences*

Le Tableau 16 ci-dessous présente la distribution des disfluences observées, après application de la méthode de recherche présentée en section 2. Quelques explications sont nécessaires pour une meilleure lecture de ce tableau :

- Colonne « @mot@ » : regroupe les résultats relatifs aux phénomènes encadrés par le symbole @. Nous avons mis en exergue @er@ et @um@ qui correspondent à ce que l'on nomme communément des hésitations. La sous-colonne « autre » regroupe tous les autres phénomènes, notamment ce que nous avons appelé « amorces » (cf. plus haut, 2.2) ;
- Colonne « répétitions » : l'immense majorité des répétitions relevées ne sont pas disfluentes, mais correspondent à la structure d'un indicatif. Le nombre de répétitions réellement disfluents est encore approximatif : certains cas ambigus restent à confirmer ou infirmer, éventuellement avec l'aide d'un spécialiste des communications de trafic de contrôle aérien.



#### 4. Présentation générale des résultats (Corpus B)

**Tableau 16** : Distribution des phénomènes de disfluences dans le corpus B (nombre d'occurrences et pourcentage)

	Hésitations (marquées par « euh »)	@mot@				Répétitions	
		Nombre total	@er@	@um@	Autre	Nombre total	Répétitions disfluentes
Nombre d'occurrences	220	1594	1497	22	74	4095	20
Pourcentage par rapport au nombre total de mots du corpus	0,18%	1,29%	0,21%	0,02%	0,06%	3,32%	0,02%

##### 4.2.2. Fréquence des corrections

Dans le Tableau 17, nous ne présentons que les résultats obtenus grâce aux marqueurs « correction », « sorry », et « pardon ». En effet, l'immense majorité des corrections (au sens précis où nous entendons ce mot) relevées sont repérables par l'un de ces marqueurs. Une analyse plus fine des résultats obtenus avec les autres potentiels marqueurs évoqués en section 2.3 est cependant nécessaire. D'ailleurs, comme pour les résultats relatifs aux répétitions disfluentes, certaines occurrences sont à analyser éventuellement avec l'aide d'un expert de la phraséologie. Les nombres donnés dans le Tableau 17 sont donc encore susceptibles d'être modifiés.

**Tableau 17** : Distribution des marqueurs de correction dans le corpus B (nombre d'occurrences et pourcentage)

	Occurrences du mot « correction »	Occurrences du mot « pardon »	Occurrences du mot « Excusez »	Occurrences du mot « sorry »	Total
Nombre total dans l'ensemble des corrections	73 / 46,49%	29 / 18,45%	3 / 1,91%	52 / 33,12%	<b>157 / 100%</b>
Nombre d'occurrences correspondant à des corrections	41 / 45,55%	25 / 27,77%	0 / 0%	24 / 26,65%	<b>90 / 100%</b>

On suppose que certaines corrections n'ont pas été dénombrées car elles ne sont pas identifiables grâce à un marqueur linguistique. Pour mémoire, nous en avons relevé 189 occurrences dans notre travail sur le corpus A pour un total de 258 corrections, ce qui montre que ce type de correction est assez fréquent. Leur connaissance est d'autant plus importante qu'elles ne peuvent être détectées que grâce à une modélisation fine de leur contexte d'apparition, avec notamment prise en compte des énoncés antérieurs.

Il est également important de noter que toutes les occurrences de « correction », « sorry », et « pardon » ne marquent pas systématiquement une correction au sens que nous avons défini, comme on peut le voir en comparant les deux lignes du tableau.

## **5. Discussion de ces résultats**

Nous avons consacré une grande partie de ce chapitre à la présentation des résultats que nous avons observés, ainsi qu'à leur commentaire. Il nous reste cependant à les considérer dans leur globalité, et selon une perspective plus générale.

De fait, nous allons considérer trois angles d'approche différents mais complémentaires, qui guident l'ensemble de cette thèse : les points de vue linguistique, cognitif et psycholinguistique, et informatique.

### ***5.1. Point de vue linguistique***

L'un des points principaux pour lequel il est nécessaire de trouver des explications est le fait qu'il y ait si peu d'erreurs et de disfluences, du moins en comparaison avec des corpus moins contraints. Cela est étonnant non seulement en soi, mais également si l'on considère que ceci concerne également le corpus d'apprentissage. De ce fait, on pourrait s'attendre à un nombre relativement grand d'erreurs, du fait de l'inadaptation des contrôleurs en formation avec la phraséologie.

Cette observation est également corroborée par les résultats obtenus sur le corpus B, dans lequel le nombre de disfluences observées est inférieur, non seulement à ceux d'autres corpus, mais même par rapport au corpus A, alors que celui-ci appartient au même domaine de tâche ! Avant d'envisager des explications d'ordre linguistique ou cognitive, il nous faut cependant proposer quelques explications plus concrètes concernant certaines caractéristiques de ce corpus. Enfin, notre principale explication est que certains phénomènes de disfluence n'ont pas été balisés en tant que tels lors de la transcription initiale. Cela paraît d'ailleurs normal, étant donné le fait que l'étude de ces phénomènes ne constituait pas un des objectifs premiers de celle-ci. Cette hypothèse est confortée par le fait que nous avons observé lors de notre analyse de corpus un certain nombre de phénomènes qui ne sont pas marqués et balisés en tant que tels. Cela concerne notamment les amorces. De plus, la qualité audio du corpus rend plus difficile la transcription fine des phénomènes. Il n'en reste pas moins que l'énorme différence observée entre les résultats obtenus avec le corpus A et ceux avec le corpus B ne peut être expliquée uniquement par ces raisons « matérielles ». Nous y reviendrons dans la section 5.2.

Nos autres explications reposent sur des considérations d'ordre pragmatique (c'est-à-dire la prise en compte de l'influence du contexte sur l'expression linguistique). Nous avons été frappé par la (relative) similarité entre la situation considérée et celle exposée dans Oviatt (1995). L'expérience exposée dans cet article met en œuvre une évaluation de

la manière dont apparaissent les disfluences selon deux modalités de production : avec format imposé (on impose au locuteur des contraintes de formes et de contenu dans sa réponse) et sans. Contrairement à ce qui paraissait prévisible, l'auteur constate (p. 32) qu'il y a beaucoup moins de disfluences (réduction de 60 à 70% !) dans le cadre de production avec format que sans. Il attribue ce phénomène à plusieurs raisons. D'une part, au fait que les contraintes ainsi imposées diminuent le temps de parole, et donc la possibilité de produire une disfluence. D'autre part, l'autre effet des contraintes de format serait d'alléger la tâche de planification du locuteur<sup>53</sup> : là encore le risque serait que des disfluences apparaissent dans la chaîne de traitement du message produit est diminué. Nous voyons dans cette étude une confirmation supplémentaire de nos propres hypothèses explicatives sur l'influence des spécificités des dialogues de contrôle aérien sur les disfluences.

### 5.2. Point de vue cognitif et psycholinguistique

#### 5.2.1. Réflexions sur les niveaux de production auxquels interviennent les erreurs

Toutes les erreurs que nous avons relevées semblent, a priori, être des erreurs immédiates telles que les a définies Reason (*op. cit.*). La raison qui nous pousse à affirmer cela est le fait que toutes les occurrences sont remarquées très rapidement par le contrôleur ou le pseudo-pilote. Ce jugement doit néanmoins être tempéré. En effet, il se base sur les seules manifestations de surface des erreurs, étant donné qu'il s'agit de l'un des principaux objets de notre analyse. Cependant, une étude plus poussée d'autres paramètres ferait peut-être apparaître des erreurs latentes, par exemple en raison de l'inscription de valeurs (par exemple, niveau de vol) erronées sur les strips.

Par rapport à la classification de Rasmussen exposée dans le chapitre 1, section 3.1.1, on peut considérer que la plupart des erreurs que nous venons de décrire sont « knowledge-based », compte tenu de la situation d'apprentissage des exercices enregistrés. Cependant, ce point doit être modéré par le fait qu'il est difficile de se prononcer avec certitude puisqu'on n'a pu mesurer avec exactitude l'ampleur de l'influence de la situation de l'apprentissage. Si l'on raisonne en termes de temporalité de l'action, il est encore plus difficile de déterminer à quel niveau appartiennent les erreurs, étant donné qu'aucun protocole expérimental n'a été mis en place qui en permettrait l'évaluation.

Basons-nous à présent sur le modèle de Levelt présenté aussi dans le chapitre 1, section 3.1.1. Nous nous intéresserons uniquement à la manière dont les erreurs se manifestent. En effet, ce sont les manifestations qui sont révélatrices du niveau auquel intervient l'erreur. La catégorie de l'erreur intervient plutôt pour expliquer le pourquoi de l'apparition de l'erreur, indépendamment de son niveau.

---

<sup>53</sup> Oviatt se situe ici dans le cadre de modélisation de la planification défini par Levelt, que nous avons présenté dans notre premier chapitre.

En 3.2.2, nous avons distingué deux catégories d'erreurs, selon qu'elles touchent la position d'un mot dans l'énoncé, ou l'énonciation du mot lui-même. Il est évident que les erreurs appartenant à la première catégorie interviennent au niveau positionnel du modèle de Levelt. En effet, elles résultent d'une mauvaise planification, au niveau syntaxique, de la chaîne produite. Cela concerne également les cas des amorces. Elles semblent *a priori* être dues à un problème d'articulation. Mais elles sont en fait provoqués par la prise de conscience par le locuteur de son erreur de positionnement. Voici un exemple avec l'énoncé

avec ENAC euh [*sis*] Littoral 627 au revoir

dont voici la représentation phonétique :

[avɛkenakœsislitoralsisãvɛsɛtɔr(ə)vwar]

Dans cet énoncé, l'apparition du son [*sis*] (mis en italique) correspond visiblement à la prononciation du début du mot « six cent vingt-sept ». Mais le locuteur se rend compte que le ce mot doit être énoncé après « Littoral », et arrête de ce fait sa production pour la remplacer par celle de « Littoral ». L'erreur n'est par conséquent pas due à un problème d'articulation, mais à la prise de conscience du locuteur qu'il n'emploie pas le bon mot au bon endroit.

Il n'en va pas de même pour les erreurs relevant de la seconde catégorie. Celles-ci sont dues à un problème situé au dernier niveau du modèle de Levelt, celui concernant la représentation phonétique. Comment arrivons-nous à cette conclusion ? Principalement par le fait que les mots qui font l'objet de ce type d'erreurs sont les mots corrects, situés à la position adéquate de l'énoncé. Cela indique que les étapes correspondant aux trois premiers niveaux ont été correctement effectuées. Par élimination, il reste le dernier niveau. Une confirmation est apportée par la nature des erreurs, qui sont toutes phonétiques.

Cependant, le nombre très bas d'erreurs et de disfluences doit être expliqué. Nous proposons comme hypothèse la faible quantité d'items compris dans le lexique, ainsi que de patrons lexico-syntaxiques distincts. En effet, de nombreuses études montrent que, tant sur les plans cognitif que psycho-moteur, plus la quantité d'éléments à choisir dans un ensemble est important, plus le temps nécessaire pour un choix donné augmente. Cela a également un impact sur le nombre d'erreurs produites au moment du choix (autrement dit, sur l'adéquation du choix avec ce qui est nécessité par la situation). Sachant que la phraséologie est moins complexe que le français utilisé couramment, nous pensons que de ce fait elle entraîne théoriquement moins d'erreurs au moment des choix sur l'axe syntagmatique. Cela est par contre compensé, du moins pour les apprenants, par la complexité des items à manipuler.

### 5.2.2. *Stratégies en présence d'une erreur*

Dans la littérature psycholinguistique, il est courant d'étudier les stratégies de récupération des erreurs par les patients. En effet, ces stratégies consistent notamment à employer des moyens détournés pour pallier la difficulté à employer tel ou tel type de lexème. Par exemple, le patient aphasique<sup>54</sup>, s'il ne peut produire une catégorie donnée de lexème, emploiera une paraphrase ou une gestuelle spécifique.

Dans notre démarche de mise en œuvre des points de comparaison avec les études sur des sujets « sains », nous avons également abordé cet aspect des difficultés de problèmes de communication chez les contrôleurs en formation. Cependant, nous avons constaté que dans aucun cas, il n'est possible de parler de stratégies palliatives au sens défini plus haut. Cela est dû à deux principales raisons. D'une part, les personnes enregistrées ne souffrant d'aucune pathologie du langage, il ne leur est pas indispensable de recourir à ce type de stratégies. La seconde raison est encore plus déterminante : là encore, le fait que le langage employé est contraint par une stricte phraséologie. Par conséquent, il n'est guère possible pour les contrôleurs de recourir à d'autres lexèmes que ceux appropriés à une position donnée de l'énoncé. De ce fait, une étude plus poussée sur ce thème devrait être menée avec un corpus relatif à une tâche moins contraignante sur le lexique utilisé. Il est par contre intéressant de se consacrer à l'analyse de la manière dont les locuteurs réagissent lorsqu'ils se rendent compte qu'ils ont commis une erreur, et signalent celle-ci à leur interlocuteur. L'intérêt est double : les phénomènes en question sont assez voisins d'une stratégie palliative ; de plus, connaître les indications que donnent les locuteurs qui se rendent compte de leurs erreurs peut servir dans l'analyse des sujets aphasiques, par exemple pour déterminer plus aisément s'ils s'aperçoivent de leurs erreurs.

### 5.3. *Point de vue informatique*

On vient de voir que les différents résultats que nous avons présentés fournissent des informations très utiles pour une connaissance approfondie de la phraséologie, ainsi que du comportement langagier des locuteurs. Toutefois, nous n'avons pas encore traité la question selon l'une des problématiques de cette thèse : le traitement automatique de ces phénomènes.

Cependant, avant d'aborder cette problématique, il nous faut répondre à une objection qui peut nous être opposée. Comme nous l'avons vu, le taux de disfluences est particulièrement bas dans le cadre des dialogues de Contrôle Aérien. De ce fait, pourquoi chercher des méthodes de traitement automatiques pour un phénomène tellement marginal ?

---

<sup>54</sup> Comme nous l'expliquons dans le chapitre 1, l'aphasie est une atteinte de certaines aires cérébrales responsables de l'activité langagière chez l'homme.

La principale motivation est liée au caractère particulièrement critique du Contrôle Aérien, que nous avons souligné en introduction et dans le chapitre 2. La prise en compte des disfluences ne concerne certes que quelques pourcents du total de mots produits. Mais dans un domaine où un seul problème, même minime en apparence, peut entraîner des conséquences graves, tout effort pour en limiter les effets est utile. Il est avéré que des problèmes de compréhension et/ou de production peuvent avoir des répercussions sur la sécurité (cf. par exemple « Bulletin Sécurité Circulation Aérienne »<sup>55</sup>, n°28 (octobre-décembre 2003) notamment p. 6 et 8, et n°45 (janvier-février-mars 2008) p.10). Le rôle spécifique des disfluences dans ces types de problèmes n'a pas encore été étudiés à notre connaissance. Mais compte tenu notamment des ambiguïtés qu'elles peuvent provoquer, leur influence est tout à fait envisageable.

Une première approche pour un traitement automatique pourrait être, comme nous l'avons fait, l'utilisation de scripts pour rechercher les disfluences à partir de marqueurs. Mais l'étude du corpus B a montré que cette recherche « automatique » est dépendante de la manière dont le corpus a déjà été « balisé » auparavant. De même, l'analyse des deux corpus a révélé que de nombreux phénomènes, particulièrement de répétition, d'auto-corrections et d'amorces, ne sont pas repérables en tant que tels, par des marqueurs qui leur seraient spécifiques.

De tout ceci, on peut tirer deux conclusions concernant l'identification automatique de ces phénomènes :

- Il est nécessaire, autant que possible, de disposer de plusieurs types de marqueurs différents pour la reconnaissance des phénomènes de disfluences. En effet, compte tenu de leur variété de manifestation, un seul type de moyen de recherche (par exemple, marqueur lexico-syntaxique, prosodie, etc.) ne suffirait pas à assurer la robustesse d'un système automatique. De plus, il faut également considérer quel est le support dans lequel la recherche est faite : signal audio, transcription électronique, etc. Selon le cas, les types de marqueurs pouvant être utilisés différeront ;
- Compter uniquement sur une ou même plusieurs catégories de marqueurs n'est pas suffisant, et parfois pas nécessaire, pour assurer une identification automatique optimum des disfluences, pour les raisons que nous avons présentées ci-dessus. Il faut donc trouver un autre moyen de décrire ces phénomènes.

Cette autre méthode de description consiste en la modélisation des disfluences, c'est-à-dire leur représentation à un niveau plus abstrait que ceux que nous avons considérés jusqu'ici.

---

<sup>55</sup> Publication interne du Service du Contrôle Aérien.

## 6. Conclusion

Le point principal à retenir de ce chapitre est l'énorme différence qui existe entre la fréquence des disfluences et leur mode d'apparition constatées dans des corpus de contrôle de trafic aérien en comparaison avec celles de tout autre corpus : il y a beaucoup moins de disfluences lorsque la phraséologie est utilisée.

Cette différence est vraie quel que soit le type de disfluence considéré. Nous avons par contre constaté que les disfluences apparaissent plus fréquemment dans le contexte de certains éléments de la phraséologie. Nous avons classifié ces éléments, non pas en fonction de critères lexico-syntaxiques comme cela se fait usuellement dans la littérature, mais selon des critères fonctionnels. Cette démarche nous est apparue en effet plus appropriée aux caractéristiques linguistiques induites par la phraséologie. Nous avons aussi établi un lien entre l'utilisation des marqueurs lexicaux et la phraséologie qui dirige normalement la production des énoncés dans le contexte de l'ATC.

Nous avons ensuite analysé plus finement les causes de cette spécificité du corpus. Pour cela, nous avons envisagé les deux niveaux qui sous-tendent toute communication humaine : linguistique et cognitif.

Enfin, nous avons envisagé l'utilisation des différents marqueurs de disfluences dans le cadre d'un traitement automatique de celles-ci. Nous avons conclu qu'ils ne sont pas suffisants à cette fin, et qu'il était nécessaire de passer à un niveau plus abstrait d'analyse de corpus. Le chapitre qui suit est consacré à cette analyse.

# Chapitre 4

## Modélisation

### 1. Introduction

Dans le chapitre 3, nous avons vu que les seules données descriptives ne sont ni nécessaires ni suffisantes pour détecter automatiquement les disfluences. Il est donc nécessaire de disposer d'informations supplémentaires. Nous avons en effet montré que certains types de disfluences peuvent n'être accompagnés d'aucun élément lexical spécifique ; de ce fait, une détection automatique utilisant ce type de marqueurs est insuffisante.

La première problématique à traiter lorsque l'on cherche à modéliser des phénomènes est le choix du formalisme à adopter. Un très important choix théorique existe en effet, compte tenu de la multiplicité des traitements souhaités et de la nature des données à traiter<sup>56</sup> en Traitement Automatique du Langage Naturel/Traitement Automatique de la Parole.

Le choix que nous avons effectué sera exposé et motivé dans la première section de ce chapitre. Nous y détaillerons également la manière dont nous l'avons mis en œuvre sur le corpus A, décrit dans le chapitre 2. La deuxième section sera consacrée à la description détaillée des différents modèles que nous avons obtenus. Enfin, la dernière section nous permettra de discuter ces résultats : nous effectuons la comparaison avec d'autres études ayant utilisé le même type de schéma, et en envisageons les conséquences d'un point de vue informatique.

### 2. Modélisation des disfluences

Nous présentons les critères pour le choix d'un formalisme de modélisation, ainsi que la manière dont nous les avons traités. Une deuxième section est consacrée à la description des principales caractéristiques du formalisme que nous avons finalement adopté. Enfin, nous abordons la question des adaptations que nous avons été amenés à effectuer sur ce formalisme, en les justifiant.

---

<sup>56</sup> Cf. par exemple Cornuéjols *et al.* (2003)



### 2.1. Critères de choix du formalisme

Les formalismes utilisés pour modéliser le dialogue oral spontané peuvent être regroupés en deux grandes catégories : les approches syntaxiques et les approches conceptuelles.

L'approche syntaxique consiste à assigner à chaque élément lexical une catégorie lexicosyntaxique, éventuellement complétée d'informations sémantiques et/ou pragmatiques. À partir de ces catégorisations sont définies des relations et des structures de plus haut niveaux (par exemple, les syntagmes, les *chunks*<sup>57</sup>, etc.). La nature de ces relations et des groupes créés varie selon le modèle théorique sous-jacent<sup>58</sup>. Les approches conceptuelles cherchent à rendre compte du sens exprimé par les composants des énoncés considérés. Les différences entre ces approches sont basées sur le formalisme utilisé pour rendre compte du sens<sup>59</sup>.

Aucun de ces deux types d'approches ne permet de modéliser de manière appropriée les disfluences dans le contexte de la phraséologie. Nous avons, en effet, montré dans le chapitre 3, section 3.2.2, que dans un tel contexte, la manière la plus pertinente de caractériser les éléments du discours est relative à la fonction qu'ils occupent dans la phraséologie, et non par rapport à la syntaxe. L'approche conceptuelle serait plus proche de la perspective fonctionnelle que nous avons ainsi adoptée dans le chapitre 3. Cependant son défaut est de ne pas rendre compte des disfluences, qui par définition ne sont pas porteuses de sens (comme nous l'avons expliqué dans le chapitre 1, section 3.2.4.6, cette affirmation est à modérer si l'on considère les disfluences comme des marques d'élaboration du discours.). De plus, comme le remarque Kurdi (p. 141), ces approches ne permettent pas de distinguer les nuances entre disfluences (par exemple entre répétition et auto-corrections).

De ce fait, il nous est nécessaire de trouver un formalisme qui corresponde aux spécificités du contexte de notre travail. Plus précisément, nous avons spécifié quatre critères principaux auxquels doit répondre le formalisme utilisé. Ils sont notamment dictés par les objectifs recherchés et le corpus à traiter. En tenant compte de ces paramètres, nous avons défini les critères suivants :

- Adaptation à l'objet d'étude : l'adaptabilité d'un formalisme est nécessaire afin de ne pas avoir à recourir à de trop nombreux ajustements *ad hoc*, qui alourdiraient le formalisme et en modifieraient l'essence. C'est l'une des raisons

---

<sup>57</sup> Abney (1991), utilisé notamment dans Goulian (2002).

<sup>58</sup> Par exemple Bove *et al.* (2006), *Construction Grammar* et *Grammaire de Propriétés* (Guénot (2005)), *Grammaire Sémantique d'Association d'Arbres* (Kurdi (2003)), analyse syntaxique de surface (Goulian (2002)), ou encore le *Fenêtrage Syntaxique* de D. Luzzati (Luzzati, 2004).

<sup>59</sup> Notamment les *structures de traits* (Pieraccini *et al.* (1995), Minker (1999)), les *Segments Conceptuels* (Pérennou (1996), Bousquet (2002)).

pour lesquelles les formalismes syntaxiques permettant de décrire l'écrit ne sont pas utilisés tels quels pour rendre compte de la syntaxe de l'oral ;

- Capacité d'expression et de représentation : le formalisme doit permettre de représenter de la manière la plus exhaustive les phénomènes étudiés, dans leur diversité ;
- Modularisation selon les besoins de l'utilisateur : si nécessaire, le formalisme doit pouvoir être adaptable par l'utilisateur, sans que ses principes fondamentaux en soient altérés.

Nous avons identifié un formalisme correspondant à ce cahier des charges dans les travaux de Shriberg *et al.* (1992), Bear *et al.* (1992) et Bear *et al.* (1993), également repris par Kurdi (2003, particulièrement p.72-75) et Heeman *et al.* (1994). Ce formalisme répond aux différentes spécifications que nous avons dressées plus haut pour un langage de représentation : il est adapté à l'objet d'étude (les disfluences), avec une grande capacité d'expression et de représentation, et enfin une modularisation selon les besoins de la modélisation (comme nous le montrons plus bas, nous avons facilement pu le modifier et l'adapter). Nous le décrivons dans la section suivante.

## **2.2. Le patron d'annotation de Bear *et al.***

### *2.2.1. Fondements théoriques*

La méthodologie générale et les principes sur lesquels le patron est fondé sont décrits en détails dans Bear *et al.* (1993).

Le but initial de ce patron est de pouvoir rendre compte de certains phénomènes de l'oral spontané, par exemple dans des transcriptions, en les étiquetant. Les auteurs désignent les phénomènes en question par le terme générique de *repairs* (en français, réparation ou correction). Comme ce terme regroupe plus de phénomènes qu'il ne pourrait le laisser penser, descendons un peu plus dans la finesse de description. Chez les auteurs, le terme de *repairs* correspond à quatre types de phénomènes. Nous les désignons selon la terminologie que nous avons nous-mêmes adoptée dans ce mémoire, avec indication des termes anglais équivalents :

- Les auto-corrections manifestées par un remplacement ou une substitution (« *repairs involving replacements (...) or insertions* ») ;
- Les répétitions d'un ou plusieurs mots<sup>60</sup> ;
- Les faux départs (« *fresh starts* ») ;

---

<sup>60</sup> On voit ici que les *repairs* ne désignent pas seulement des corrections, puisque les répétitions, bien que disfluentes, sont également désignées par ce terme.

- Tout phénomène mettant en action des fragments de mots (« *cases involving a word fragment* »). Au vu des exemples donnés et de la description, les auteurs ont la même définition que nous des fragments de mots.

Même si ce n'est pas dit explicitement, le modèle théorique sous-jacent est celui de Shriberg (1992)<sup>61</sup> que nous avons déjà présenté en détails dans le premier chapitre. De ce fait, toute zone dans laquelle apparaît un « *repair* » est partagée en trois zones : le *Reparandum*, l'*Interruption Point*, et le *Repair*. Le patron d'annotation consiste à rendre compte de la position et de la nature des éléments situés dans cette zone. Pour cela, les auteurs ont défini un ensemble de symboles et de règles pour les utiliser, que nous présentons ci-après.

### 2.2.2. Méthodologie d'annotation

La méthodologie de traitement des disfluences est locale : le but n'est pas de rendre compte de la totalité de l'énoncé, mais uniquement du contexte pertinent pour représenter la disfluence considérée et son environnement immédiat.

Ce contexte est composé de trois parties. Le pivot est la *zone d'édition* (ou *Interruption Point*, que nous dénommerons dans la suite *point d'édition*), représentée par une barre verticale « | ». A gauche et à droite de celle-ci se situent respectivement le *Reparandum* et la zone de correction (*Repair*). C'est dans le cadre de l'ensemble constitué par ces deux zones que sont annotés les phénomènes. A partir du symbole de la zone d'édition sont dérivées deux autres annotations de la zone d'édition, en fonction du contexte dans lequel se situe celle-ci : « .| » lorsqu'il y a une relation sémantique entre le *Reparandum* et la zone de correction ; « || » lorsqu'il est difficile d'établir une relation entre les deux zones. De fait, les auteurs utilisent la plupart du temps la seule barre verticale. Nous faisons de même.

Pour l'annotation elle-même, il est fait recours à cinq symboles différents, chacun correspondant à un type d'élément (en général, des mots) relatif au phénomène à annoter :

- M pour « *matching word* » (autrement dit « *mot correspondant* »). Comme son nom l'indique, il s'agit ici de rendre compte des cas où le même élément se trouve de part et d'autre de la zone d'édition. Lorsque ce symbole apparaît plusieurs fois dans un même énoncé, chaque occurrence est associée à un index numérique qui permet de la distinguer des autres. Exemple<sup>62</sup> :

---

<sup>61</sup> Qui est l'un des co-auteurs de l'article.

<sup>62</sup> Chaque exemple donné ici est repris de Bear *et al.* (1992). A chaque fois nous donnons les pages correspondantes.

*I'd like I'd like to stop in Washington*

M1 M2 .| M1 M2 (p. 3)

- R pour « *replacement* » (« *remplacement* » en français). Il s'agit de tout élément situé dans le *reparandum*, et faisant l'objet d'un remplacement dans la zone de correction. Le remplacement peut être effectué par un mot appartenant à la même catégorie syntaxique et/ou morphologique que le mot remplacé. Comme pour le symbole « M », et selon les mêmes conditions, un indice numérique est appliqué à ce symbole. Exemples<sup>63</sup> :

*to the city at Atlanta in Atlanta using ground transportation*

R1 M1 | R1 M1

*What are the cheap cheapest one way flights*

R1 | R1 (p.4)

- X : Ce symbole n'est pas l'abréviation d'un mot. Il sert à marquer les éléments qui font l'objet d'une suppression (présents dans le *Reparandum*, absents dans la zone de correction) ou d'une addition (absents dans le *Reparandum*, présents dans la zone de correction). Exemple (ici, une addition) :

*List the aircraft list types of aircraft ...*

M1 X M2 .| M1 X X M2 (p. 5)

- C (pour « *Cues* »). (« *indices* » en français). Désigne tous les marqueurs tels que « désolé », « non », etc. qui marquent la réaction du locuteur à une disfluence apparue dans sa production. Exemple :

*from Atlanta back to Pittsburgh I'm sorry back to Denver*

M1 M2 R1 C C | M1 M2 R2 (p. 5)

- FP (pour *Filled Pauses*). Il s'agit ici de noter toute « pause remplie » telle que « euh » par exemple. Exemple :

*How long is the layover in Denver uh in Dallas*

M1 R1 FP | M1 R1 (p. 6)

Il n'y a pas de symbole spécifique pour représenter les faux départs. La raison est que Bear *et al.* préfèrent utiliser le tiret (« - ») ; celui-ci est utilisé à la manière d'un diacritique, en suffixation du symbole principal correspondant à un faux-départ. Exemples :

<sup>63</sup> Pour faciliter la compréhension des exemples, chaque élément de l'énoncé faisant partie du patron est mis en italique, et le symbole correspondant du patron est placé immédiatement en dessous.

Also list *fl-* *flights* from Atlanta to Boston...

M1- | M1 (p. 7)

*on July fif-* *on July twentieth*

M1 M2 R1- | M1 M2 R1 (p. 6)

Le marquage d'un faux-départ peut se combiner avec n'importe lequel des symboles présentés ci-dessus, à l'exception évidemment de celui correspondant au « pauses pleines » (FP). Par définition, on ne peut parler à leur égard de « faux départ », puisque ce ne sont pas des unités lexicales porteuses de sens.

Enfin la dernière question traitée par les auteurs est celle de l'étendue de la zone à annoter. Il est en effet exclu de prendre en compte l'ensemble de l'énoncé comportant la disfluence, mais uniquement celle-ci et son contexte. La réponse de Bear *et al.* varie en fonction du type de disfluence modélisé et de la position par rapport au point d'édition (gauche ou droite). Pour les auto-corrections, ne seront annotés que les éléments qui font l'objet d'une suppression dans le *Repair* (côté droit) par rapport au *Reparandum* (côté gauche).

### 2.3. Ajouts et modifications apportés au patron de Bear *et al.*

En analysant les caractéristiques de ce patron, on observe qu'il est principalement axé sur la modélisation des disfluences qui entraînent une perturbation de l'énoncé (par exemple auto-corrections, amorces, répétitions, etc.). Par ses effets, ce type de disfluences s'oppose ainsi à celles qui ne sont qu'une insertion de phénomènes éditeurs, non-porteurs de sens, comme les hésitations, les pauses longues, etc. La raison est que les hésitations nécessitent d'être représentées de façon abstraite : il s'agit d'un phénomène ponctuel, qui, à part son occurrence, ne perturbe pas plus que ça l'organisation de l'énoncé. De fait, les hésitations et autres seront traitées indirectement de par leur incorporation dans les patrons. Nous adoptons nous aussi cette perspective de traitement des disfluences.

Cependant, il nous a été nécessaire d'apporter, à d'autres niveaux, quelques modifications au patron que nous venons de décrire dans la section précédente. En effet, ce patron est très généraliste et est initialement plus destiné à des tâches de transcription et d'annotation<sup>64</sup> que de modélisation informatique.

Les modifications que nous avons apportées concernent essentiellement la mise en place de nouveaux symboles. Cela répond à deux objectifs. Le premier est de rendre compte de manière plus précise de la nature des différentes catégories de disfluences. Nous avons en

---

<sup>64</sup> Signalons toutefois que les auteurs affirment (p. 1) qu'il peut également servir à des tâches de détection automatique. Mais cette possibilité n'est plus évoquée par la suite, contrairement à d'autres applications.

effet jugé préférable de conserver la mention de la nature des disfluences dans leur représentation. C'est pour la même raison que nous avons distingué les marqueurs exprimant la correction et ceux exprimant l'excuse, alors qu'ils sont regroupés sous la catégorie générale « *Cues* » chez Bear *et al.* Notre second objectif est d'adapter le patron d'annotation à certaines spécificités de la tâche et du corpus.

Le Tableau 18 ci-dessous décrit les symboles introduits et leur explication. On voit qu'ils correspondent pour la plupart aux marqueurs linguistiques que nous avons identifiés dans le chapitre précédent.

**Tableau 18 :** Adaptation du patron de Bear *et al.* à notre corpus

Symbole ajouté	Phénomènes représentés par le symbole
<b>Corr</b>	Tout mot ou groupe de mots exprimant explicitement une correction, tel que « correction », « je corrige », etc.
<b>Exc</b>	L'expression d'une excuse, comme « désolé », « sorry » ...
<b>Brib</b>	Correspond à ce que nous avons appelé « fragment de mot ».
<b>PV</b>	Une pause longue, au sens défini dans le chapitre 2 (d'une durée supérieure à 0,2 seconde).
<b>Neg</b>	L'expression d'une négation par un mot ou groupe de mots : « non », « ce n'est pas », etc.
<b>^</b>	<p>Ce diacritique est employé par Bear <i>et al.</i> pour rendre compte des contractions telles que <i>I will</i> → <i>I'll</i>, qui sera représenté par <i>M1 M2</i> → <i>M1^M2</i><sup>65</sup>. Dans les productions dictées par la phraséologie, nous n'avons pas repéré ce type de contractions liées à la langue anglaise.</p> <p>Par contre, nous avons « recyclé » ce symbole pour représenter les éventuels changements de prononciations des séquences de lettres et ou de chiffres utilisés dans les indicatifs, les fréquences, etc. (de part et d'autre du point d'édition ; par exemple : <i>contact ENAC 120 contact ENAC 1 2 6</i> sera représenté par le patron <i>M1 R1 / M1 M2 R1^R2^R3</i>).</p> <p>Comme le montre cet exemple, la règle d'annotation que nous avons adoptée est de représenter par un seul symbole une séquence alphanumérique prononcée comme une seule unité (ici, <i>120</i>, prononcé « cent vingt ») et par plusieurs symboles agrégés par le caractère « ^ », une prononciation en plusieurs éléments (ici, <i>1 2 6</i>, prononcé « un deux six » au lieu de « cent vingt six »).</p>

<sup>65</sup> Il n'est employé que rarement c'est pourquoi nous ne l'avons pas mentionné plus haut.

Symbole ajouté	Phénomènes représentés par le symbole
-	Il ne s'agit pas d'un ajout au sens strict ; par rapport à Bear <i>et al.</i> , nous avons choisi d'utiliser ce symbole, en plus de son usage sur les symboles M et R, en combinaison avec X. Nous avons en effet considéré que certaines des suppressions constatées dans le corpus correspondent également à des amorces. Dans ce cas, nous avons jugé utile de le représenter dans les patrons, dans la perspective d'identifier ce qui fait l'objet de la correction.

Nous avons également traité un cas potentiellement problématique : l'occurrence de plusieurs corrections les unes à la suite des autres, dans un même énoncé. Ce cas n'est pas abordé dans Bear *et al.* Le problème, ici, est de déterminer si l'on représentera cette suite dans un seul et unique patron, ou chaque correction par un patron distinct. Nous avons choisi cette deuxième solution. La raison de ce choix est que chaque occurrence de correction concerne un élément distinct de celui corrigé par les autres corrections. Cela pourrait certes apparaître comme un biais potentiel. On pourrait en effet nous objecter qu'il est possible que ces occurrences soient liées d'une manière ou d'une autre. Cependant, cela répond à un impératif méthodologique qui est de dégager des patrons à partir des différentes occurrences considérées séparément. D'autre part, le nombre de cas où plusieurs auto-corrections ou amorces se suivent dans un même énoncé est très faible. Notre choix n'affectera donc pas l'interprétation générale des phénomènes, ni leur traitement.

Enfin, en ce qui concerne la question de l'étendue de la zone d'annotation, nous calquons notre méthodologie sur celle de Bear *et al.*

### 3. Description des patrons observés

Dans cette partie, nous présentons les différents patrons que nous avons relevés dans l'ensemble du corpus.

Dans un premier temps, nous donnons un aperçu général des résultats les plus notables. Dans les deux sous-sections qui suivent, nous présentons les résultats nécessitant plus d'analyses et de commentaires. Ces sections seront différenciées selon le trait « Patrons sans amorces » VS « Patrons avec amorces ». Cette distinction nous est apparue nécessaire pour des considérations théoriques et fonctionnelles. Théoriques, car la catégorie des patrons sans amorce ne concerne que les auto-corrections, alors que les patrons avec amorces peuvent correspondre à d'autres phénomènes. Et fonctionnelles, puisque les amorces ont un mode de fonctionnement spécifique.

Chaque section sera elle-même divisée en sous-parties, afin de rendre compte des catégories suivantes :

- Patrons « simples », c'est-à-dire ne comprenant aucun autre symbole que ceux correspondant à la similitude et/ou le remplacement (respectivement « M » et « R »). Nous les appelons « simples » car du point de vue des traitements automatiques qui peuvent en être dérivés, ce sont les patrons qui posent le moins de difficultés ;
- Patrons comportant une ou plusieurs additions, déplacements, substitutions, omissions. Nous donnons plus de détails dans la section 3.1 ci-dessous ;
- Autres patrons.

Cette classification d'analyse est la plus pertinente dans notre contexte, pour plusieurs raisons. D'abord, elle permet d'appréhender clairement les modifications introduites par les occurrences distinctes de ce que nous avons nommé les « patrons simples ». D'autre part — et c'est un corollaire du point précédent —, il sera éventuellement possible de cibler ainsi, en fonction de ces catégories, les différents traitements pouvant être accomplis.

### ***3.1. Aperçu général***

Notre but est, dans un premier temps, de disposer d'une représentation générale de la distribution des patrons de disfluences et d'en dégager les principaux traits distinctifs. Nous avons effectué cette modélisation à partir du corpus A, et en prenant en compte les ajouts que nous avons présentés en section 2.3.

Nous reprenons dans le Tableau 19 la catégorisation, déjà utilisée dans le chapitre 3, en Addition, Déplacement, Omission, et Substitution. Seulement, alors que nous l'avions alors employée au niveau des éléments lexicaux, nous l'utilisons à un niveau plus large, afin de caractériser les patrons en fonction des différences entre les éléments situés de part et d'autre du point d'édition. Ainsi, l'addition désigne l'insertion, à droite du point d'édition, d'un ou plusieurs éléments lexicaux absents de la partie gauche ; le déplacement concerne le changement de position, dans la partie droite par rapport à celle de gauche, d'éléments présents des deux côtés ; l'omission est la suppression d'un ou plusieurs éléments lexicaux situé à gauche du point d'édition mais absent à droite ; enfin la substitution est le remplacement d'un ou plusieurs éléments lexicaux de la gauche par un ou plusieurs éléments lexicaux à droite.

Ces quatre catégories sont croisées avec les différentes catégories fonctionnelles établies dans le chapitre 3. Pour chacune de celles-ci, nous avons effectué une sous-catégorisation afin d'indiquer si les corrections se font avec une reprise d'un ou plusieurs éléments situés avant le point d'édition (« reprise ») ou que la correction s'opère



immédiatement au niveau de l'élément corrigé. Nous indiquons le nombre d'occurrences et le pourcentage correspondant (séparés par un « / »).

On observe que les substitutions sont de loin la modalité de correction la plus fréquente, et également celle dont la distribution est la plus uniforme, quelles que soient les fonctions du discours concernées. A l'inverse, on n'observe que très peu d'omissions, et elles apparaissent toutes dans le contexte d'indicatifs.

**Tableau 19 :** Distribution des principaux types de patrons d’auto-correction selon les catégories fonctionnelles

	Indicatifs		Niveaux et autres équivalents (directions, caps, etc.)		Fréquences		Ordres, requêtes, question, etc.		Erreurs de langue		Politesse	Pourcentage
	Reprise	A partir de l'élément corrigé	Reprise	A partir de l'élément corrigé	Reprise	A partir de l'élément corrigé	Reprise	A partir de l'élément corrigé	Reprise	A partir de l'élément corrigé		
Addition	1 / 15%	5 / <b>15%</b>	4 / <b>12%</b>	5 / <b>15%</b>	0	0	0	6 / <b>18%</b>	0	0	13 / <b>38%</b>	<b>15%</b>
Déplacement	0	3 / <b>60%</b>	0	0	0	0	0	1 / <b>20%</b>	0	0	1 / <b>20%</b>	<b>2%</b>
Omission	6 / <b>100 %</b>	0	0	0	0	0	0	0	0	0	0	<b>3%</b>
Substitution	13 / <b>7%</b>	37 / <b>21%</b>	13 / <b>7%</b>	48 / <b>27%</b>	0	9 / <b>5%</b>	5 / <b>3%</b>	33 / <b>19%</b>	1 / <b>1%</b>	15 / <b>8%</b>	3 / <b>2%</b>	<b>80%</b>

Un problème se pose pour la représentation en patron des déplacements, car rien n'est prévu pour cela dans le schéma de Bear *et al.* Nous pallions ce problème en les considérant comme des cas de remplacement. Voici par exemple comment nous modélisons le déplacement marqué en italique dans l'énoncé suivant :

**Lufthansa 7 6 euh 6 7 9**

**R1 R2 FP | R1 R2**

Notons enfin que, dans la suite de ce chapitre, nous regroupons les substitutions dans la catégorie des patrons simples. En effet, selon la définition des substitutions que nous avons donnée plus haut, il s'agit de remplacements, qui appartiennent donc à cette catégorie de patrons.

### 3.2. Patrons sans amorces

Dans les sous-sections suivantes, nous présentons sous la forme de tableaux les différents types de patrons sans amorces que nous avons observés. Nous présentons d'abord les patrons simples, puis les patrons correspondant aux quatre catégories présentées en 3.1. Nous concluons par les patrons de répétitions disfluentes, que nous distinguons des patrons simples car ces derniers ne concernent que les auto-corrections.

Pour chaque patron, nous donnons le nombre d'occurrences relevées, ainsi qu'un exemple d'énoncé correspondant, extrait du corpus d'étude. Dans l'exemple, la partie en gras correspond à la zone représentée par le patron. Lorsqu'il n'y a qu'une seule occurrence d'un patron, nous faisons également figurer l'énoncé correspondant dans le tableau, en indiquant qu'il s'agit de la seule occurrence et non d'un exemple parmi d'autres.

#### 3.2.1. Patrons simples

**Tableau 20 :** Composition et distribution des patrons simples (sans amorce)

Patrons et exemples	Fréquence (en nombre d'occurrences)
<i>Un mot directement changé par un autre</i>	
R1   R1	
<u>Exemple</u> : France Air Force <b>40 44</b> 30 contactez ENAC 123 décimale 8	9

Patrons et exemples	Fréquence (en nombre d'occurrences)
<i>Un mot directement changé par un autre, avec au moins une hésitation ou une pause vide dans le patron<sup>66</sup></i>	
R1 FP   R1	7
R1 PV   R1	4
<i>Deux mots changés par deux autres</i>	
R1 R2   R1 R2  <u>Occurrence</u> : Luxair 7 3 5 euh <b>descendez niveau descend level</b> euh 7 0	6
<i>Deux mots changés par deux autres, avec au moins une hésitation dans le patron</i>	
R1 R2 FP FP   R1 R2	1
<i>Un mot changé par un autre, dans un syntagme</i>	
M1 R1   M1 R1  Exemple : niveau 1 2 0 <b>route</b> <b>Lacan route Amboise</b>	22
R1 M1   R1 M1  <u>Exemple</u> : 41 1000 feet euh above us <b>crossing for descending for your</b> level	9
R1   R1^R2^R3  <u>Exemple</u> : contact ENAC euh <b>1 2 9</b> <b>129</b> 4	2
R1 M1 M2 M3 M4   R1 M1 M2 M3 M4  <u>Occurrence</u> : Speebird <b>7 4 4 ENAC</b> <b>bonjour</b> [oe] 5 4 4 ENAC bonjour radar identified [oe] climb level 1 2 0 route Amboise Lacan Châteaudun .	1
<i>Un mot changé par un autre, dans un syntagme avec au moins une hésitation dans le patron</i>	

<sup>66</sup> Nous ne donnons pas d'exemples car la seule différence avec le patron précédent est la présence d'une hésitation ou d'une pause longue.

Patrons et exemples	Fréquence (en nombre d'occurrences)
M1 R1   FP M1 R1	7
R1 M1 FP   R1 M1	1

On observe que deux patrons se dégagent par leur fréquence. D'une part, le changement immédiat d'un mot par un autre, avec reprise d'un seul élément situé avant ou après le mot corrigé (31 occurrences au total).

D'autre part, le remplacement immédiat d'un mot par un autre, sans intervention d'aucun autre élément dans le contexte (20 occurrences au total)

Enfin, on observe l'occurrence d'une structure plus complexe. Son analyse permet de voir que la correction d'une erreur — et donc, *a priori*, sinon sa détection du moins sa prise en compte dans la boucle de production— peut intervenir jusqu'à un empan de 5 mots à partir de sa production. Nous disons « apparemment » car, bien que ce soit la mesure de ce type la plus grande que nous ayons relevé dans notre corpus, nous n'excluons pas que ce nombre puisse être plus grand encore. Dans tous les cas de figure, il s'agit d'un type d'occurrence assez rare.

### 3.2.2. Patrons comportant une ou plusieurs additions

Les patrons de ce type sont plus rares : un total de 34 occurrences (15% du total), soit autant que la principale catégorie de patrons simples. On observe également une plus grande disparité dans les configurations exprimées par les patrons. Il peut y avoir, selon les occurrences, de 1 à 6 éléments qui sont ajoutés. Les additions peuvent apparaître à n'importe quelle position au sein du *Repair* ; de fait, dans la majorité des cas, elles se situent immédiatement après le point d'édition, et avant la répétition d'un élément situé dans le *Reparandum*. Ces occurrences indiquent que le locuteur se rend compte qu'il a oublié un ou plusieurs éléments situés avant le mot qu'il vient de prononcer. On voit que dans la majorité des cas, le locuteur remarque son erreur quasi-immédiatement (après la prononciation d'un seul mot).

**Tableau 21 :** Composition et distribution des patrons comportant au moins une addition (sans amorce)

Patrons et exemples	Fréquence (en nombre d'occurrences)
<i>Addition d'un seul élément après le point d'édition</i>	
M1   X1 M1	11
<u>Exemple</u> : D R O bonjour <b>level</b> <b>maintain</b> level 1	

Patrons et exemples	Fréquence (en nombre d'occurrences)
6 0	
<i>Addition de plusieurs éléments après le point d'édition</i>	
$M1 M2 M_i   X1 X2 \dots Xn M1$ (où $i \in [1 - 3]$ et $n \in [1 - 6]$ ) <sup>67</sup> <u>Exemple 1</u> : Luxair euh 7 3 4 <b>maintain</b> ENAC <b>good morning maintain</b> level 1 5 0 <u>Exemple 2</u> : F K T descendez <b>niveau</b> ENAC <b>bonjour et descendez niveau</b> 110	5
<i>Cas particuliers (pas plus de deux occurrences dans tout le corpus)</i>	
$M1 M2   M1 X1 M2$ <u>Occurrence 1</u> : ENAC euh good morning DLT <b>0 2 0 6: 2</b> euh level 1 5 0 route Lacan <u>Occurrence 2</u> : on a passé <b>le 160 le niveau 160</b> en montée France Air Force 69 69	2
$R1 M1 FP M2   R1 X1 M1 M2$ <u>Occurrence</u> : euh radar identified euh control service provided from euh Balon euh <b>climb level euh 1 join: Balon level 1 0 0</b> route euh Balon Limoges euh report Balon	1
$R1 M1 M2   X1 R1 M1 M2$ <u>Occurrence</u> : à Kingair 1000 pieds en <b>dessous à travers votre montez en travers votre</b> niveau .	1
$R1 FP   FP X1 X2 FP PV R1$ <u>Occurrence</u> : <b>requesting euh euh which level euh request ?</b>	1

### 3.2.3. Patrons comportant une ou plusieurs omissions

Etant donné le très faible nombre d'omissions (un total de trois sur l'ensemble du corpus !) il est difficile de dégager des patrons communs. Malgré tout, on notera la variabilité de chaque occurrence, que ce soit en termes de positionnement par rapport au point d'édition et aux autres éléments du patron, ou de nombre d'éléments supprimés lors de la correction.

<sup>67</sup> Rappelons que les valeurs de  $i$  et de  $n$  données ici sont observées sur le corpus. Sur un corpus différent, ces valeurs peuvent changer, comme nous le montrons en section 4 ci-dessous.

**Tableau 22 :** Composition et distribution des patrons comportant au moins une omission (sans amorce)

Patrons et exemples	Fréquence (en nombre d'occurrences)
M1 M2 M3 X1   M1 M2 M3 <u>Exemple</u> : euh Luxair 7 <b>3 5 4 7 3 5</b> euh correction maintain level 1 1 0 .	1
M1 FP X1 M2 FP PV   M1 M2 <u>Exemple</u> : euh P H L F ENAC good morning euh climb level euh 1 1 [oe] 0 euh euh <b>route euh Poitiers</b> <b>Cognac euh route Cognac</b> [cont] euh	1
X1 X2 R1 FP FP   R1 <u>Exemple</u> : F B T E E ENAC bonjour maintenez niveau 100 euh route euh Poitiers puis direct Saumure <b>je</b> <b>vous rappelle euh euh rappelez</b> Poitiers pardon	1

### 3.2.4. Autres patrons

Comme pour les omissions, le très petit nombre d'occurrences empêche d'aboutir à des conclusions fiables.

**Tableau 23 :** Composition et distribution des autres catégories de patrons sans amorce

Patrons et exemples	Fréquence (en nombre d'occurrences)
R1 R2 R3 FP   R1^R2^R3 <u>Occurrence</u> : euh TAT euh 289 M euh I contact ENAC euh <b>1 2 9 euh 129</b> 4.	1
R1^R2^R3 FP   R1 <u>Occurrence</u> : euh D C <b>MAX euh euh M A euh</b> <b>X</b> ENAC good afternoon maintain level 1 6 0 route Poitiers Amboise euh Tours report Amboise.	1

### 3.2.5. Patrons de répétitions

La principale observation que l'on peut faire en analysant les patrons figurant dans le Tableau 24 rejoint celle que nous faisons dans le chapitre 3 sur les répétitions disfluentes : d'un point de vue formel également, ces disfluences présentent une structure de formation très peu complexe (1 à 2 éléments répétés maximum, pas de combinaisons complexes entre les éléments compris dans le patron). On remarque également que près de la moitié des

occurrences de répétitions disfluentes concernent des éléments non porteurs de sens : les hésitations. Nous interprétons cette observation comme la preuve d'une vigilance accrue des contrôleurs aériens et pilotes dans le cadre de leur production orale : lorsqu'une hésitation est produite, ils sont déjà dans un mode moins attentif, et sont par conséquent plus vulnérables à d'autres types de disfluences.

**Tableau 24 :** Composition et distribution des patrons de répétitions disfluentes

Patrons et exemples	Fréquence (en nombre d'occurrences)
<p>C1   C1</p> <p><u>Exemple</u> : F Q B plan de vol clôturé à 11 heures 38 <b>euh euh</b> vous pouvez quitter la fréquence au revoir</p>	10
<p>M1   M1</p> <p><u>Exemple</u> : F euh E T montez niveau 1 8 0 euh et <b>maintenez maintenez</b> vitesse euh 200 200 Kts euh maximum</p>	10
<p>M1 M2   M1 M2</p> <p><u>Exemple</u> : France Air Force 66 10 euh <b>cause trafic cause trafic</b> euh maintenez niveau 1 6 0 .</p>	2
<p>C1 M1   C1 M1</p> <p><u>Occurrence</u> : <b>euh Speedbird euh Speedbird</b> 5 4 4 euh climb level 1 4 0 euh i call you back for higher .</p>	1

### 3.3. Patrons avec amorces

Nous présentons les résultats obtenus dans des tableaux, selon les mêmes conventions que dans la section 3.2. Par contre, nous ne donnons pas d'énoncés à titre d'illustration, car la plupart des exemples donnés pour les patrons sans amorces sont valables ici aussi.

#### 3.3.1. Patrons simples

Nous avons fait l'hypothèse que les patrons spécifiques aux amorces, simples ou non, seraient moins nombreux que ceux relatifs aux cas qui n'en comportent pas, étant donné que l'analyse linguistique nous en avait montré la faible fréquence par rapport aux autres. Par contre, un fait est étonnant. On pouvait s'attendre a priori qu'ils présentent plus ou moins des distributions de configurations identiques à celles constatées pour les amorces de corrections ne comportant pas d'amorce, nonobstant la différence générale de



fréquence. Or il n'en est rien. Le patron le plus fréquent est relatif à la reprise d'une amorce par son équivalent complété : M1- | M1. Il ne s'agit donc pas d'une correction à proprement parler puisque le mot faisant l'objet de la reprise est le même que celui qui a été interrompu. Ce patron est donc interprétable comme une « marque de formulation du discours » et non comme une correction. On observe d'ailleurs que les variantes de ce patron (plusieurs mots repris, présence de pauses ou hésitations, etc.) sont également très fréquentes. Cette observation peut certes s'expliquer par le fait la nature même des phénomènes représentés par cette catégorie de patrons s'y prêtent. On peut quand même s'étonner que les cas d'interruption au milieu même d'un mot qui doit être corrigé ne soient pas plus nombreux.

**Tableau 25** : Composition et distribution des patrons simples avec amorce non-corrective

Patrons et exemples	Fréquence
M1-   M1	9
<i>Patrons avec au moins une hésitation ou pause vide</i>	
M1- FP   M1	6
M1- PV   M1	1
M1- M2   M1 M2	1
M1 M2 M3-   M1 M2 M3	1
M1 M2 M3 M4 M5-   M1 M2 M3 M4 M5	1

**Tableau 26** : Composition et distribution des patrons simples avec amorce corrective

Patrons et exemples	Fréquence
R1-   R1	4
M1 R1-   M1 R1	3
R1^R2-   R1	1
M1 M2 R1-   M1 M2 R1	1
<i>Patrons avec au moins une hésitation</i>	
R1- FP   R1	7
R1 M1- FP   R1 M1	1

En complément de cette observation, on remarque également que les patrons ne correspondant pas à une correction sont légèrement majoritaires : 19 occurrences pour 17 occurrences relatives à une correction.

La deuxième constatation à faire dans ce tableau est une plus grande variété dans la formation des patrons, avec par contre une fréquence plus faible pour chacun. Nous n'avons pas d'hypothèse particulière pour l'expliquer.

### 3.3.2. *Patrons d'amorces comportant une ou plusieurs additions ou omissions*

Compte tenu du fait qu'il y a très peu d'amorces comportant des additions ou des omissions, nous les regroupons dans la même section et le même tableau.

**Tableau 27** : Composition et distribution des patrons d'addition ou omission avec amorce corrective

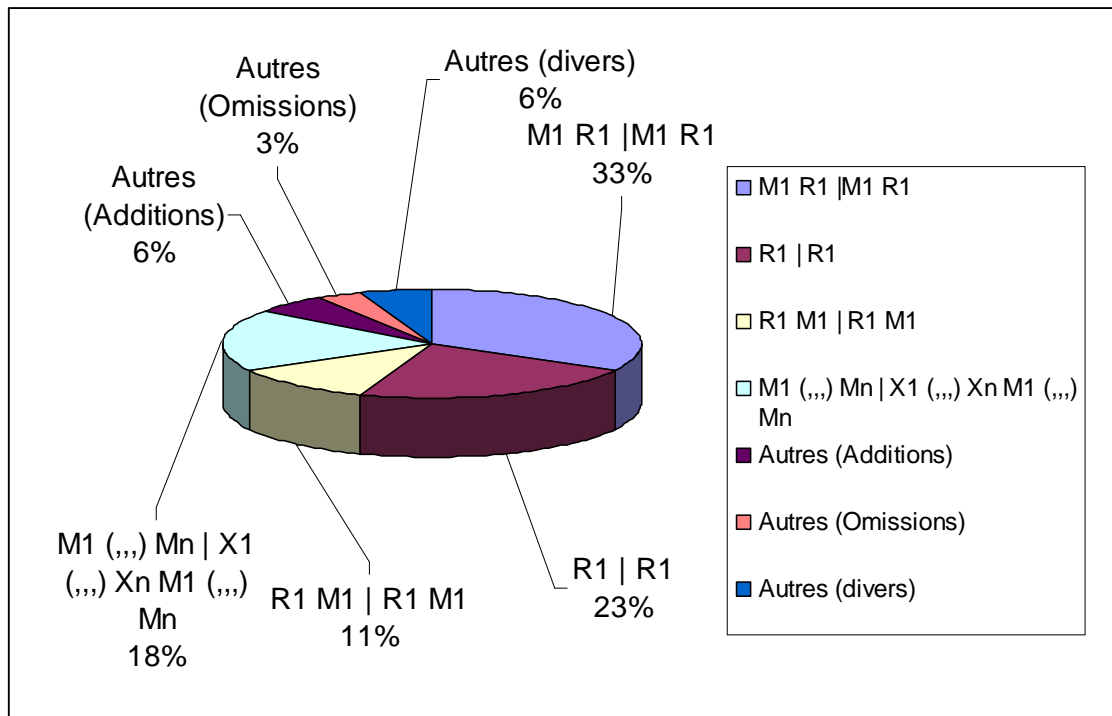
Patrons et exemples	Fréquence (en nombre d'occurrences)
<i>Addition d'un seul élément après le point d'édition</i>	
R1-   X1 R1	3
<i>Addition de plusieurs éléments après le point d'édition</i>	
R1-   X1 X2 X3 R1	2
R1 M1- PV   X1 FP R1 M1	1
R1   X1 X2 X3 X4 X5 X6 R1	1
<i>Omission d'un ou plusieurs éléments après le point d'édition</i>	
X1 FP X2 X3- FP   X1 X2	1
X1- FP   X1 X2	1
M1 X1 X2 M2   M1 M2	1
M1 FP M2 X1 M2 M3- FP   M1 M2 M3	1

Le nombre total d'occurrences est trop faible pour pouvoir tirer des conclusions assurées. Cependant, on remarque une fois encore la différence de manifestation des additions/omissions lorsqu'elles se manifestent dans le contexte d'une amorce.

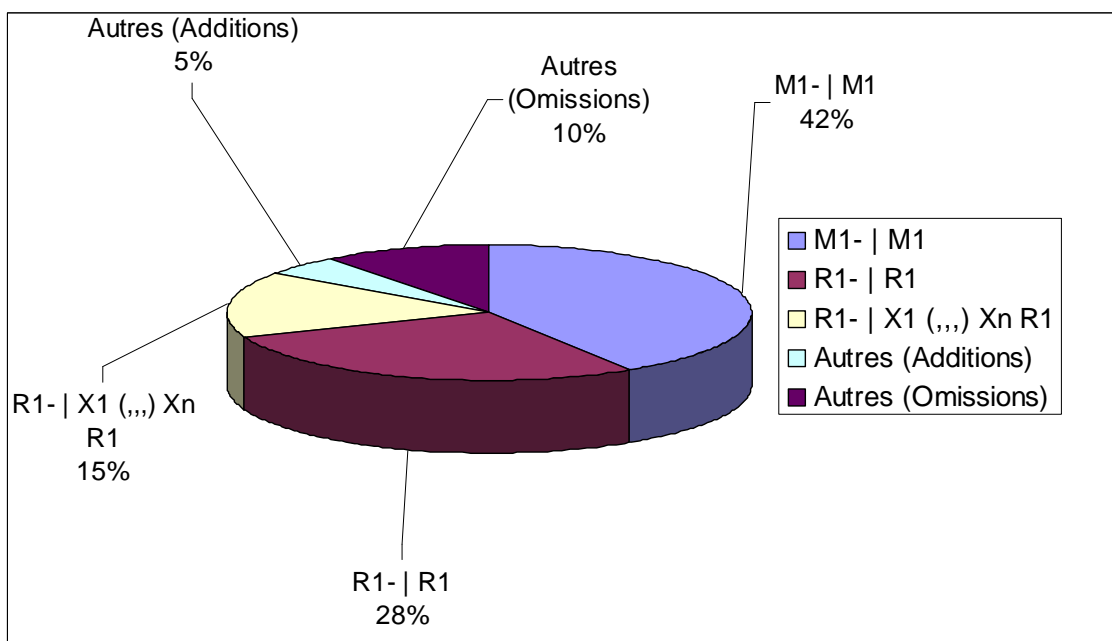
Ainsi, l'apparition des additions semble être liée à certains contextes privilégiés. Pour les amorces, toujours dans le contexte d'une substitution. Et nous avons vu que pour les patrons sans amorce, c'est dans le contexte d'une répétition que les additions apparaissent le plus fréquemment.

#### 4. Discussion

Pour faciliter l'analyse des principaux résultats, voici deux représentations synthétiques, en secteurs, de la distribution des différents patrons. Nous avons regroupé les patrons présentant des similarités. La Figure 13 est consacrée aux patrons sans amorce, la Figure 14 aux patrons avec amorce.



**Figure 13 :** Synthèse de la distribution des principaux patrons de correction sans amorce



**Figure 14 :** Synthèse de la distribution des principaux patrons d'amorces (correctives ou non)

Ces figures nous permettent d'illustrer les conclusions que nous pouvons tirer de la mise en place et de l'analyse des différents patrons. Trois principales observations ressortent, concernant la nature des patrons :

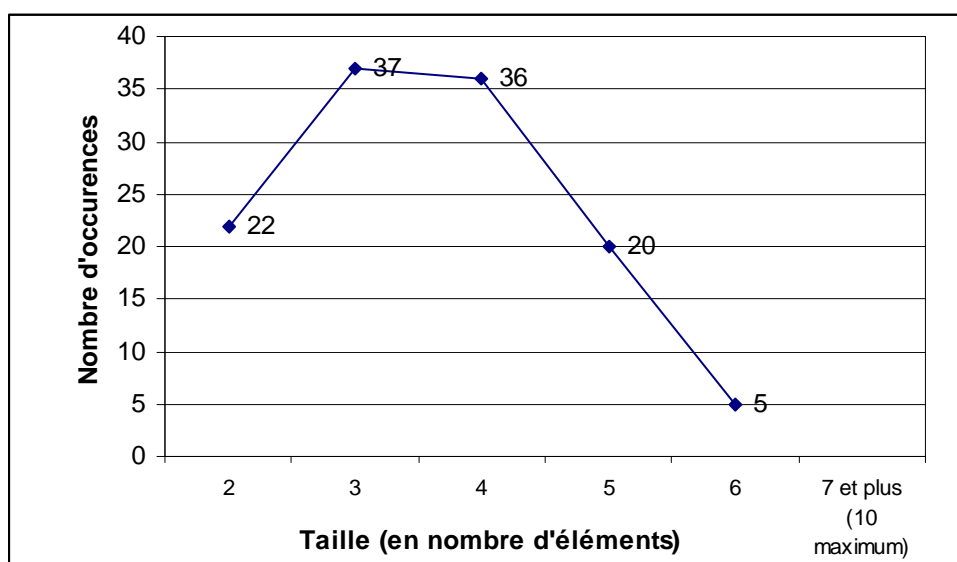
- Leur variété (nombre de patrons différents entre eux d'au moins une unité) n'est pas très importante. Il n'y a en effet que 4 principales catégories de patrons (pour les corrections, 3 pour les amorces). Les autres patrons, souvent une seule occurrence, représentant 15% du total (aussi bien pour les corrections que pour les amorces). Comparons avec les études similaires mais menées sur des corpus différents. Kurdi (*op. cit.*, p. 72) recense un total de 48 patrons différents (toutes disfluences confondues). Heeman *et al.* (1994, p. 297) dénombrent<sup>68</sup> 72 patrons différents pour les seuls phénomènes de correction (sans prise en compte des patrons incluant des fragments et des termes éditeurs). Enfin, il n'y a pas de décompte des patrons dans Bear *et al.* (1992, 1993) et Shriberg *et al.* (1994) ; on remarquera cependant que le nombre d'éléments consécutifs pouvant être omis peut dépasser 6 (pour 3% du total : Bear *et al.* (1992, p. 57))<sup>69</sup>. Comme nous l'avons montré, dans notre corpus le nombre le plus élevé d'éléments faisant l'objet d'une addition ou d'une omission est de 6 ;
- *Leur taille est assez petite* : par « taille », nous entendons le nombre total d'éléments (représentés chacun par un symbole) d'un patron. La Figure 15<sup>70</sup> montre la corrélation entre la fréquence d'un patron et sa taille. On voit que les tailles associées aux plus grandes fréquences sont de 3 et 4 éléments ; les tailles immédiatement inférieures et supérieures (respectivement 3 et 5) font l'objet de fréquences moindres ; au-delà de 5 éléments, la fréquence diminue très rapidement. Nous interprétons cette observation de la même manière que nous l'avons fait pour la majorité de patrons de type R1|R1 :

---

<sup>68</sup> Sur un corpus de dialogues entre deux personnes collaborant pour accomplir une tâche : l'une cherche à résoudre le problème, l'autre l'assiste.

<sup>69</sup> Le corpus d'étude est composé de dialogues de planification de voyages en avion, selon un protocole de « Magicien d'Oz » (le locuteur croit parler à un système de dialogue automatique, qui est en fait simulé par un opérateur humain).

<sup>70</sup> Conçue à partir de l'ensemble des patrons considérés dans ce chapitre, y compris ceux relatifs aux amorces non-correctives.



**Figure 15 :** Rapport entre la taille des patrons et leur fréquence

le locuteur se rend compte rapidement de son ou ses erreurs, et cherche également à se corriger le plus vite possible. Cette explication est corroborée par le fait que plus de 67% des patrons de taille supérieure à 2 sont des patrons de type R1|R1, M1 R1| M1 R1 ou R1 M1 |R1 M1, dans lesquels apparaissent parfois des éléments n'ayant pas de rapport direct avec la correction effectuée (par exemple, hésitation, excuse, négation, etc.). Ce résultat est cependant différent de celui relaté par Kurdi :

**« La tendance observée dans les répétitions est aussi confirmée avec les autocorrections : la fréquence d'un patron est inversement proportionnelle à sa taille. » (Kurdi, *op. cit.*, p. 76).**

- Ils sont, en majorité, simples. Nous employons le terme de « simplicité » dans le même sens que celui que nous avons défini en début de section 3. Cette caractéristique découle de la précédente, puisqu'un patron d'une taille réduite ne se prête pas à des combinaisons complexes, et que la rapidité du locuteur à se corriger limite également l'intrication des éléments.

Tous ces résultats ne sont pas étonnants compte tenu de la nature du corpus et de nos analyses précédentes. Ils sont dans la continuité des résultats obtenus lors de l'analyse des marqueurs des disfluences.

Enfin, si l'on se place du point de vue du traitement informatique des disfluences, il est évident que ces caractéristiques auront un impact sur le choix des algorithmes de traitement, comme nous le montrerons au chapitre 5.

## **5. Conclusion**

Dans ce chapitre, nous avons cherché à obtenir une représentation abstraite des disfluences de notre corpus. Plus précisément, nous avons modélisé chaque occurrence trouvée, au moyen du patron d'annotation proposé par Bear *et al.* (1992), patron que nous avons adapté par certains aspects.

Nous avons ainsi pu dégager un ensemble de caractéristiques formelles des disfluences, que ce soit en termes de taille, de positionnement. Nous avons également montré que les spécificités de la tâche de contrôle aérien impliquent certaines particularités dans la forme des patrons relevés, par rapport à celles observées dans des études sur d'autres types de corpus. On constate en effet une moins grande diversité des patrons, ainsi qu'une complexité moindre.

Maintenant que la phase d'analyse du corpus est achevée, nous disposons d'une connaissance précise des modes de manifestation des disfluences dans notre corpus. Aussi bien du point de vue lexico-syntaxique que d'un point de vue plus abstrait. Nous pouvons à présent utiliser ces données pour un traitement informatique des disfluences.



**PARTIE 3**

**TRAITEMENT AUTOMATIQUE DES**

**DISFLUENCES**





## Chapitre 5

# Vers un traitement automatique des disfluences

### 1. Introduction

Le but de ce chapitre est d'envisager la manière dont nous pouvons traiter automatiquement ou semi-automatiquement les phénomènes de disfluences.

Dans un premier temps, nous étudierons les principales méthodes qui ont été utilisées par des travaux antérieurs. Ces méthodes ne sont toutefois pas utilisables telles quelles. En effet, comme nous l'avons vu dans les deux chapitres précédents, les disfluences présentent plusieurs spécificités lorsqu'elles apparaissent dans le contexte du contrôle de trafic aérien.

Nous chercherons donc, dans un deuxième temps, à mettre en place une méthodologie de détection automatique qui prenne en compte ces spécificités. Pour cela, nous utiliserons à la fois les résultats des travaux cités dans la première section, et les observations que nous avons faites sur notre corpus. Pour respecter l'équilibre entre les chapitres, nous resterons à une granularité de description assez large : nous décrirons les principes que nous comptons appliquer, mais non les modalités d'application elles-mêmes, que nous réservons à un chapitre ultérieur.

### 2. Approches existantes

Lorsqu'on l'envisage d'un point de vue informatique, l'expression orale spontanée comporte deux composantes. D'une part l'aspect *traitement du signal*. Il s'agit du traitement du son qui a été produit. Les méthodes employées utilisent les paramètres classiques décrits en phonétique : durée du signal, intensité, force, etc. Les différentes caractéristiques du signal sont analysées selon différentes méthodes mathématiques (coefficients spectraux, Modèles de Markov Cachés, etc.)<sup>71</sup>. Au final, le but est de décoder le signal, de manière à obtenir sa décomposition en unités significantes (les phonèmes). A partir de cette suite d'unités, le moteur de traitement du signal reconstitue des unités plus larges : mots, syntagmes, phrases ...

C'est en sortie du moteur de traitement du signal qu'intervient la deuxième composante, qu'on désigne habituellement par *compréhension de la parole*. Il s'agit d'interpréter la signification d'ensembles d'éléments lexicaux (habituellement un ou plusieurs énoncés)

---

<sup>71</sup> Pour un panorama de ces méthodes, cf. Calliope (1989).

qui ont été produits par le système de traitement du signal. La tâche est difficile, d'autant plus qu'en général le système de traitement du signal fournit, pour un énoncé donné, plusieurs énoncés différents pouvant correspondre à ce qui a été effectivement produit (on parle des «  $N$  meilleures solutions », l'adjectif « meilleur » indiquant qu'il s'agit des solutions les plus probables, classées habituellement par ordre décroissant de probabilité).

Nous suivons cette distinction bipartite pour organiser notre présentation des différentes méthodes de traitement automatique des disfluences : selon qu'elles interviennent au niveau du traitement du signal, ou à celui de la compréhension de la parole.

### 2.1. Niveau du traitement du signal

Les travaux basés spécifiquement sur le traitement du signal lui-même sont assez rares. Cela est normal étant donné la difficulté pour traiter le signal sans erreurs ou avec le minimum d'erreurs. Malgré les évolutions de la technologie de traitement du signal, il y a encore beaucoup de progrès à faire.

Il existe d'ailleurs un consensus implicite dans la littérature sur cette quasi-impossibilité de reconnaître les disfluences uniquement grâce à des indices acoustiques<sup>72</sup>. On en trouve une formulation claire dans Bear *et al.* (1992):

**“While acoustics alone cannot tackle the problem of locating repairs, since any prosodic patterns found in repairs are likely to be found in fluent speech, acoustic information can be quite effective when combined with other sources of information, in particular with pattern matching.” (Bear *et al.*, 1992, p. 60)**

Cette constatation, bien que datant de plusieurs années, est encore vraie de nos jours, malgré les progrès scientifiques et techniques dans les techniques de reconnaissance et de traitement du signal. On lit ainsi dans Shriberg *et al.* (2004) :

**“Most problems for which prosody is a plausible knowledge source can be cast as statistical classification problems. By that we mean that some linguistic unit  $U$  (e.g., words or utterances) is to be classified as one of several target classes  $S$ . The role of prosody is to provide us with a set of features  $F$  that can help predict  $S$ .” (Shriberg *et al.*, 2004, p. 105)**

Elle est également corrélée par le fait que très peu de travaux de notre bibliographie cherchent à utiliser uniquement les caractéristiques acoustiques des disfluences. Et lorsque c'est le cas, l'utilisation pour la détection automatique des disfluences<sup>73</sup> est évoquée, mais

---

<sup>72</sup> Parmi les indices acoustiques, c'est essentiellement la prosodie (que nous avons définie rapidement dans le chapitre 1, section 2.1.4) qui est utilisée.

<sup>73</sup> C'est notamment le cas de Nakatani *et al.* (1994), Plauché *et al.* (1999), Oviatt (1995), Shriberg *et al.* (1992).

non implémentée. Cette nuance est importante à préciser : l'analyse et l'interprétation sont alors effectuées par un opérateur humain, même lorsqu'il est assisté par des outils logiciels.

Les autres études utilisant effectivement des particularités acoustiques pour le repérage des disfluences, n'en font usage que comme une source supplémentaire d'information et de discrimination, ajoutée à d'autres (par exemple lexicales).

Il est enfin à signaler qu'une disfluence en particulier pose particulièrement problème au niveau de la reconnaissance de la parole. Il s'agit de ce que nous dénommons, dans notre terminologie, les amorces (correctives ou non). Le problème résulte du fait que la majorité des moteurs de reconnaissance de la parole fournissent en sortie des mots qui font partie du vocabulaire qui a servi pour leur apprentissage. De ce fait, toute suite de phonèmes absente du vocabulaire du système sera ignorée ou éventuellement ramenée au mot le plus proche existant. Par exemple, si le système doit décoder l'amorce « joi- », et que son vocabulaire comporte les mots « joie », « joindre », et « jointure », il pourra faire correspondre l'un de ces trois mots à la chaîne phonétique qu'il reçoit en entrée<sup>74</sup>. Il pourra éventuellement produire ainsi une solution assez satisfaisante si le mot correspond à l'amorce. Ce cas de figure n'est pourtant qu'un pis-aller : d'une part il est aléatoire ; d'autre part, il ne rend pas compte des cas où l'amorce est corrective ; enfin, il ne permet pas d'identifier l'amorce en tant que telle.

Ce problème est évoqué par différents auteurs. Shriberg (1994, p. 26) l'évoque sans proposer de solution. Nakatani *et al.* (1994, p. 1610-1611) envisagent deux solutions possibles. D'une part traiter les amorces comme des occurrences d'un seul et unique élément générique du modèle de langage du système. Mais les auteurs remarquent à juste titre que cette proposition ne rend pas compte de la diversité des occurrences d'amorces<sup>75</sup>. De ce fait, la deuxième solution proposée consiste à mener une analyse « *bottom-up* ». Par rapport à la première solution, cette dernière aurait l'avantage de pouvoir identifier directement les phonèmes appartenant aux amorces. Par contre, l'idée est seulement décrite théoriquement. Plus récemment, Liu (2003) établit plusieurs propriétés acoustiques et prosodiques caractéristiques des amorces. Mais comme l'auteur le remarque elle-même, ce travail, bien que prometteur, est encore préliminaire. De plus, la remarque que nous faisons dans la note de bas de page 75 est valable également pour son approche.

## 2.2. Niveau de la compréhension de la parole

Le niveau de la compréhension de la parole se situe en aval de la reconnaissance de signal. Il ne s'agit donc plus de traiter du signal mais des éléments lexicaux

---

<sup>74</sup> Pour un exemple tiré d'un cas réel de transcription automatique d'un corpus (*Switchboard*), voir Liu (2003, p. 37).

<sup>75</sup> Nous ajoutons que cela ne permet également pas d'identifier le mot que le locuteur voulait prononcer, et donc si il s'agit d'une amorce auto-corrective ou non.

(éventuellement d'autres éléments selon le paramétrage du module de traitement du signal : par exemple des marques d'hésitation). Nous avons recensé plusieurs approches différentes pour le traitement des disfluences au niveau de la compréhension de la parole.

Pour en simplifier la présentation, nous les regroupons sous trois catégories :

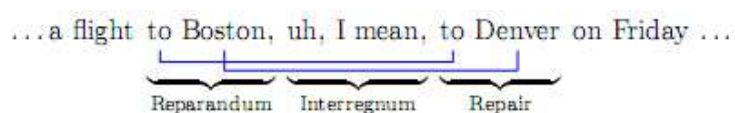
- Statistiques et probabilistes ;
- Syntaxico-sémantiques ;
- Patrons.

qui sont présentées dans les sous-sections éponymes. Les différences entre ces catégories ne sont évidemment pas tranchées. En effet, les auteurs emploient souvent des combinaisons de plusieurs d'entre elles. Dans ce cas de figure, nous présentons chaque aspect du travail considéré dans la catégorie correspondante.

### 2.2.1. Approches statistiques et probabilistes

Les méthodes probabilistes les plus souvent utilisées dans la compréhension automatique de l'oral spontané sont les Modèles de Markov Cachés (en anglais : *Hidden Markov Model*, abrégé en *HMM*). En effet, ils permettent une certaine souplesse dans le traitement. Sans entrer dans les détails formels, rappelons qu'ils permettent de modéliser et prédire la probabilité d'apparition d'un phénomène discret (par exemple un élément lexical) en fonction des  $N$  phénomènes précédents.

C'est en raison de leur mode de fonctionnement que les *HMM* sont rarement utilisés en tant que tels pour le traitement des disfluences. Comme le font remarquer Johnson *et al.* (2004b, p. 33), ce type de formalisme est difficilement applicable pour le traitement de disfluences. En effet, il repose sur une conception linéaire de la chaîne des éléments de l'énoncé. Or, cette conception n'est pas utilisable pour certaines disfluences telles que les auto-corrections parce que les différentes parties impliquées dans ce type de disfluences<sup>76</sup> présentent ce que les auteurs nomment des « dépendances croisées ». Pour comprendre ce concept, voici l'exemple tiré de leur article. Soit l'auto-correction représentée dans la Figure 16 :



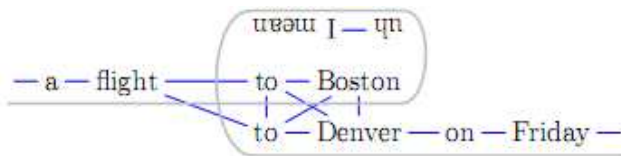
**Figure 16 :** Un exemple d'auto-correction, avec représentation des différentes zones (tiré de Johnson *et al.* (2004b, p. 34)).

Si l'on représente l'écoulement du temps pendant l'énonciation de l'énoncé par une ligne, ainsi que les dépendances croisées, on aboutit à la représentation graphique de la

---

<sup>76</sup> Les auteurs se basent sur la représentation de Shriberg en *Reparandum*, *Interregnum* et *Repair*.

Figure 17. Y apparaissent clairement les relations entre le mot précédent le début de l'auto-correction ("*flight*") et les deux composants de celle-ci, ainsi que les relations existant entre eux. Cette figure montre d'une manière intuitive qu'un traitement uniquement linéaire de la disfluence ne suffit pas<sup>77</sup>.



**Figure 17 :** Représentation graphique des dépendances entre les zones de l'auto-correction de la Figure 16 (tiré de Johnson *et al.* (op. cit.))

Les auteurs proposent donc d'utiliser un formalisme plus adapté, car moins dépendant, pour l'analyse d'un énoncé, du contexte précédant l'élément courant. Il s'agit des Grammaires à Arbres Adjoints (en anglais : *Tree Adjoining Grammars*, soit *TAG*). Ce formalisme a plus sa place dans la section 2.2.2 ci-dessus. En effet, l'utilisation qui en est faite est à base de syntaxe. Nous renvoyons donc à cette section pour une description plus détaillée.

L'aspect plus proprement statistique utilisé par Johnson *et al.* (2004b) (et développé dans d'autres articles des mêmes auteurs : Charniak *et al.* (2001), Johnson *et al.* (2004a), Lease *et al.* (2006)) repose sur l'idée qu'à un énoncé disfluent donné, correspond un ou plusieurs énoncés fluents (c'est-à-dire dépourvus de toute disfluence) possible. Le but est alors de trouver l'énoncé le plus probable. Le modèle de disfluence utilisé est celui de Shriberg, que nous avons présenté dans le premier chapitre, avec la distinction entre les trois zones : *Reparandum*, *Interregnum*, et *Repair* (comme présenté dans la Figure 16).

Comme dans la majorité des autres études sur les disfluences, trois principaux paramètres sont utilisés pour évaluer ces systèmes<sup>78</sup> :

- Taux de rappel : nombre de disfluences correctement détectées par rapport au nombre total d'occurrences ;
- Taux de précision : nombre de détections correctes par rapport au nombre total de détections (y compris les « faux positifs ») ;
- Taux d'erreur : pourcentage d'erreurs dans l'exécution d'une tâche. En l'occurrence, la tâche considérée consiste à identifier une disfluence ou un élément appartenant à une zone disfluente.

<sup>77</sup> Cette analyse s'applique également aux *Context Free Grammars* (CFG) et leur version probabilisée (PCFG). Nous n'en détaillons pas la nature dans ces pages, car nous les décrivons dans le chapitre 6, dans un autre contexte.

<sup>78</sup> Cf. définitions données par exemple dans Johnson *et al.* (2004b, p. 38) ou Heeman *et al.* (1994, p. 296).

Les résultats rapportés dépendent du type de disfluence faisant l'objet de l'étude. Pour la détection des mots appartenant au *Reparandum*, Johnson *et al.* (2004b, p. 38) rapportent des taux de précision allant de 77% à 82%, et des taux de rappel allant de 76% à 77% (selon les paramètres utilisés). Dans un autre article (Johnson *et al.* (2004a)), les mêmes auteurs cherchent à extraire trois types de phénomènes : les mots situés dans le *Reparandum*, les mots faisant partie d'un segment d'édition (hésitations, « so anyway », etc.), et le point d'*Interruption*. Selon la modalité d'entrée (signal audio ou transcription) et le type de phénomène à identifier, le taux d'erreur varie entre 76% pour le plus élevé (reconnaissance de mots du *Reparandum*, sur du signal) à 28,60% pour le plus faible (reconnaissance du point d'*Interruption* sur des transcriptions ; Johnson *et al.* (2004a, p. 6))

Une autre approche statistique pour le traitement des disfluences est employée par Stolcke *et al.* (1996). Comme pour l'approche précédente, elle repose sur une hypothèse principale : la prise en compte des disfluences d'un énoncé donné est plus robuste si elle se base sur l'énoncé correspondant, dépourvu de toute disfluence (Stolcke *et al.*, *op. cit.*, p. 405). En revanche, les disfluences sont traitées comme des « événements », au même titre que des mots. Ainsi, elles peuvent être traitées de manière séquentielle par un HMM<sup>79</sup> (cf. p. 406 *et sq.*). Le but des auteurs est d'évaluer l'amélioration de la robustesse d'un modèle de langage par la prise en compte de certaines disfluences. Les auteurs montrent que le taux d'erreur reste quasi-identique, avec ou non prise en compte des disfluences (respectivement, 50,23% et 20,21% ; cf. Stolcke *et al.*, *op. cit.*, p. 406). Dans la suite de leur article, les auteurs analysent les raisons de ce résultat mitigé. Cette analyse est plus orientée vers l'amélioration des modèles de langage en général que vers la reconnaissance des disfluences, qui nous intéresse ici.

Une troisième approche est utilisée dans Heeman *et al.* (1994). Elle est basée sur l'usage de certains types de disfluences (par exemple, hésitations, termes d'édition, etc.) comme indices des occurrences d'auto-corrections. Pour cela, les auteurs mettent en place un modèle statistique. Il associe à la probabilité d'occurrence d'une disfluence donnée, la probabilité d'apparition d'une auto-correction (Heeman *et al.* (1994, p. 300)). Les auteurs indiquent, pour la détection de disfluence sur le corpus de test, un taux de rappel de 83% et un taux de précision de 89% à l'aide de cette méthode. Ils précisent cependant que la comparaison avec d'autres résultats est délicate, en raison notamment des différences de typologie des disfluences (Heeman *et al.* *op. cit.*, p. 301).

---

<sup>79</sup> Ceci n'est pas contradictoire avec les affirmations de Johnson *et al.* (2004b). En effet, le modèle proposé par Stolcke *et al.* (1996) ne modélise que des disfluences pouvant être représentées ainsi, à savoir les pauses remplies, les répétitions, et les additions de mots. Les auteurs confirment eux-mêmes (p. 405) qu'ils n'intègrent pas dans ce modèle des disfluences plus complexes telles que les substitutions.

### 2.2.2. Approches syntaxico-sémantiques

Nous avons déjà rencontré dans le chapitre 4 certains des formalismes syntaxico-sémantiques utilisés aussi bien pour la modélisation que pour le traitement automatique des disfluences. Rappelons le principe de ce type d'approches. Elles sont basées sur des représentations syntaxique et / ou sémantique des énoncés traités. Par syntaxe, on entend la description des relations des constituants entre eux, par exemple au moyen des « parties du discours » auxquels ils appartiennent. Ces relations définissent des unités plus larges (par exemple les syntagmes) auxquelles sont rattachés les constituants. Ces unités entretiennent elles-mêmes des relations entre elles, et peuvent former des unités de plus haut niveau (l'unité maximale étant l'énoncé).

On observe une grande variété de formalismes utilisés selon les auteurs<sup>80</sup>. Ils diffèrent notamment selon le mode de représentation des relations syntaxiques, et l'utilisation ou non de la sémantique. Notre but n'étant pas de faire ici un catalogue exhaustif de toutes les approches existantes, nous n'en mentionnerons que deux à titre d'illustration.

Commençons par la *Grammaire d'Arbres Adjoints*, ou TAG, déjà évoquée plus haut. Elle est utilisée, selon des variantes différentes, dans les différentes déclinaisons des travaux de Charniak *et al.* (2001), ainsi que par Kurdi (2003). Appliquée à la compréhension automatique du langage, la TAG consiste à représenter les constituants de l'énoncé traité sous la forme d'une suite « d'arbres ». Chaque arbre rattache un ou plusieurs des constituants à un élément plus générique. La constitution des arbres est dictée par des règles de bonne formation. Différentes opérations peuvent être effectuées sur les arbres pour rendre compte des relations entre chacun : adjonction, disjonction, etc.

Ce formalisme est adapté au traitement des disfluences du fait de la souplesse d'analyse qu'il permet par rapport à des approches plus rigides. En effet, il n'impose pas un sens de traitement défini, contrairement aux HMM. De plus, il permet de rendre compte des dépendances croisées que nous évoquons plus haut, puisque les règles de formation des arbres permettent de rattacher les constituants aux mêmes arbres, selon différents critères. Kurdi (2003) propose quant à lui une version alternative de la TAG, la *Sm-TAG* (pour *Semantic Tree Association Grammar*). Elle est basée sur une combinaison entre la *grammaire sémantique* définie par F. Rastier et la *LTAG* (*Lexicalized TAG*). Ce formalisme est lui-même une dérivation de la TAG initiale. Son principe est d'utiliser les catégories lexico-syntaxiques pour caractériser les nœuds des arbres. Il s'agit ensuite de rattacher les disfluences à des arbres de plus haut niveau. Cette approche, associée à une recherche par

---

<sup>80</sup> Signalons cependant que là encore, la plupart des travaux relevés utilisent ces formalismes pour représenter les disfluences, mais n'évoquent pas, ou à peine, leur implémentation.

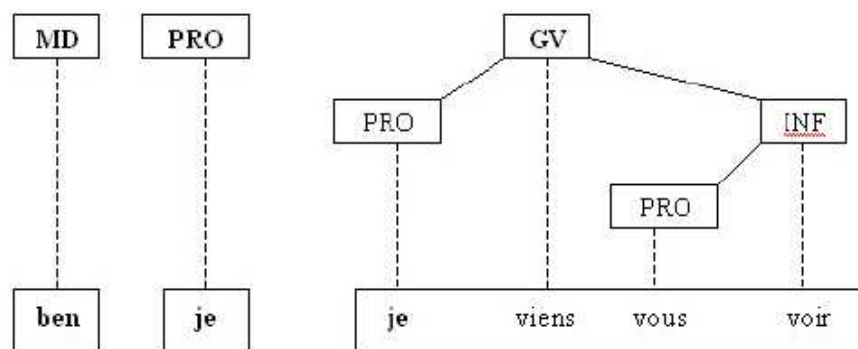


patrons (cf. 2.2.3 ci-dessous) permet d'obtenir, en détection, un taux de rappel de 89,67% et un taux de précision de 92,76% (Kurdi (2003, p. 171))<sup>81</sup>.

Une méthode similaire est proposée par Bove *et al.* (2006). Elle s'inscrit dans le cadre d'un système existant de traitement de la parole<sup>82</sup>, dans lequel les énoncés produits sont représentés sous la forme d'arbres syntaxiques<sup>83</sup>. Elle consiste à rattacher les constituants disfluents aux nœuds adéquats de la représentation. Ainsi<sup>84</sup>, l'énoncé

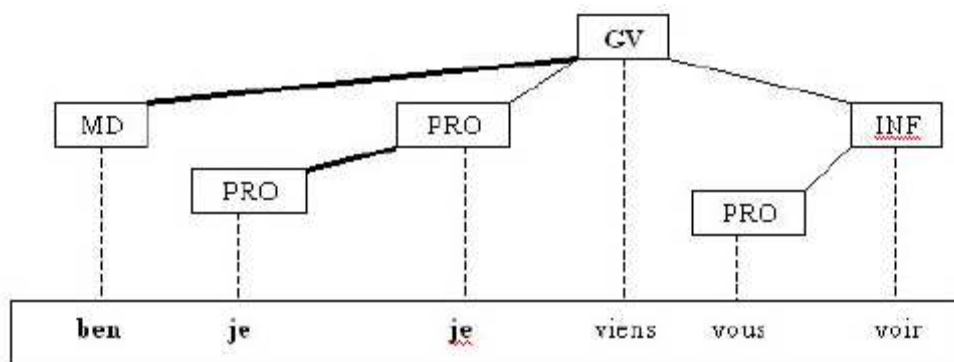
ben **je** je viens vous voir

sera représentée, avant traitement, comme dans la Figure 18 :



**Figure 18 :** Représentation syntaxique d'un énoncé disfluent avant traitement spécifique (système TiLT)

Après prise en compte des disfluences, la représentation obtenue est la suivante (Figure 19 ci-dessous).



**Figure 19 :** Représentation syntaxique d'un énoncé disfluent après traitement spécifique (système TiLT)

<sup>81</sup> L'auteur évalue trois systèmes différents, et donne beaucoup plus de résultats dans son travail. Nous ne mentionnons ici que ceux qui nous paraissent les plus parlants.

<sup>82</sup> Système TiLT, présenté dans C. Chardenon, "Analyse syntaxique en dépendances et Evaluation". In : *Traitement Automatique des Langues Naturelles (TALN 2005), Atelier Evaluation*, volume 2, p.25, Dourdan, mai 2005.

<sup>83</sup> Formés selon une grammaire de dépendance.

<sup>84</sup> L'énoncé, ainsi que les deux figures qui suivent, sont tirés de Bove *et al.* (2006, p. 108). Dans l'énoncé, la disfluence est mise en gras dans le texte initial.

Pour arriver à ce résultat, plusieurs pré-traitements sont nécessaires. En premier lieu, un « enrichissement » du lexique afin de rendre compte des marqueurs de disfluences (tels que « ben » dans l'exemple donné). D'autre part, la mise en œuvre de nouvelles règles dans le moteur de traitement syntaxique. Leur rôle est de définir les conditions de rattachement à un nœud de l'arbre syntaxique, ainsi que le résultat obtenu après rattachement (par exemple, fusion des éléments répétés en un seul). Enfin, un « contrôle sémantique » est ajouté pour optimiser l'action de ces règles. Il consiste à faire correspondre à chaque élément du lexique une catégorie sémantique ; à partir de cette catégorisation un algorithme est défini, qui permet de sélectionner le candidat le plus pertinent parmi l'ensemble des arbres possibles pour un énoncé donné. L'article ne faisant pas état d'une évaluation, on ne sait pas quels sont les résultats de cette approche.

### 2.2.3. *Approches par patrons*

L'approche par patrons est utilisée, à notre connaissance, dans au moins trois travaux différents (Bear *et al.* (1993), Heeman *et al.* (1994), et Kurdi (2003, description aux pages 147-153)). Elle repose sur une modélisation préalable des disfluences sous la forme d'occurrences de patrons, à partir des travaux de Bear *et al.* (1992). Nous l'avons nous-mêmes utilisée dans le chapitre précédent.

Le traitement consiste à utiliser des algorithmes qui vont rechercher la présence de patrons caractéristiques d'un ou plusieurs phénomènes de disfluence. Les patrons recherchés ont été préalablement établis par les chercheurs au moyen d'études sur des corpus similaires à ceux sur lesquels est effectué le traitement.

L'avantage de cette approche est principalement une grande simplicité de mise en œuvre, puisque les algorithmes de détection sont relativement simples à concevoir. De plus, elle est généralisable à d'autres types de corpus ou de tâches.

Par contre, cette méthode à elle seule ne rend pas compte de toutes les subtilités de manifestation des disfluences dans une production orale. Elle produit notamment des « faux positifs », autrement dit des détections erronées de phénomènes non disfluents. Pour cette raison, elle est complétée par l'utilisation, soit d'une autre approche (telle que celles citées précédemment dans ce chapitre), soit par des sources d'information complémentaire. Cette dernière possibilité consiste à conserver comme approche principale la recherche des patrons, mais à utiliser, en plus des seuls éléments lexicaux traités, des informations sur leur catégorie syntaxique ou le sens qu'ils véhiculent. Cela passe évidemment par un enrichissement des données fournies en entrée du traitement automatique.

Par exemple, Bear *et al.* (1992) font appel au moteur d'étiquetage syntaxique et sémantique *Gemini*. Ces auteurs rapportent un taux de reconnaissance des « repairs » de

96% avec utilisation d'informations syntaxiques, contre 62% avec le seul moteur de reconnaissance des patrons (Bear *et al.* (1992, p. 58-59)). De même, Heeman *et al.* (1994) ont recours à un étiquetage syntaxique, couplé au modèle statistique que nous avons décrit en 2.2.1. Sans ce modèle, ils obtiennent, pour la détection sur un corpus de test, un taux de rappel de 91,5%, et un taux de précision de 45.3%. Ces résultats sont à comparer avec ceux que nous exposons en 2.2.1, notamment le gain de plus de 40% pour le taux de précision

On voit ainsi que, quelle que soit la méthode choisie pour compléter la reconnaissance de patrons, elle donne de meilleurs résultats que cette approche seule.

### **3. Approche proposée**

#### ***3.1. Adaptation au corpus et au contexte***

Les approches ayant trait au signal ne sont pas utilisables dans le contexte que nous étudions, et ce pour deux raisons. D'une part, comme on l'a vu, elles ne sont pas encore au point et servent dans la majorité des cas d'indices supplémentaires à un autre mode de détection. D'autre part, le traitement automatique du signal est vraisemblablement difficile dans le cadre des communications de contrôle aérien. En effet, la qualité acoustique du canal est loin d'être parfaite : parasites, pertes de certaines fréquences de la voix, distorsion de la voix, etc. Ainsi, le traitement automatique du signal, déjà difficile, donnerait de nombreuses erreurs en raison de tous les « bruits » du canal.

Nous postulons également qu'il ne serait pas productif d'utiliser des techniques reposant sur des statistiques. En effet, elles demandent, pour produire des résultats, une grande quantité de données d'entraînement, dont nous ne disposons pas. De plus, l'un des intérêts de la modélisation statistique est la finesse d'expression et de prédiction de phénomènes complexes. Cette caractéristique nous apparaît inutile dans le cadre qui nous intéresse ici : nous avons montré dans le chapitre 4 que les modèles considérés sont simples. Ainsi, étant donnés les inconvénients que nous venons d'évoquer, avoir recours à cette technique poserait plus de problèmes techniques et théoriques que d'avantages.

#### ***3.2. Principaux éléments de notre proposition***

Dans un premier temps, nous envisageons les différents traitements que nous pouvons appliquer aux disfluences. Nous nous basons pour cela sur les conclusions de notre état de l'art ainsi que nos réflexions précédentes sur l'adaptation à notre corpus.

Une deuxième section expose de manière plus détaillée les traitements envisagés pour les disfluences les plus complexes à prendre en compte.

### *3.2.1. Identification des traitements à appliquer*

Le premier traitement, duquel découlent tous les autres, est l'identification de l'occurrence de la disfluence. Ce seul point nécessite le choix d'une ou plusieurs méthodologies, parmi l'ensemble de celles que nous avons répertoriées dans les sections précédentes.

Compte tenu des contraintes que nous avons présentées, l'approche la plus appropriée nous paraît être l'utilisation de règles de reconnaissance de patrons. Notre raisonnement est double.

D'une part, en raisonnant par élimination, c'est la méthode présentant le moins de contraintes, compte tenu des inconvénients des autres approches, que nous avons exposés en section 3.1.

D'autre part, cette approche est selon nous celle qui permet le mieux de rendre compte du mode de formation des manifestations des disfluences.

Enfin, elle est adaptable, moyennant quelques ajouts, à tous les types de disfluences que nous avons recensés dans notre chapitre 3. Ce n'est pas le cas de toutes les approches, par exemple avec l'utilisation des HMM, comme nous l'avons vu (cf. 2.2.1 ci-dessus).

Enfin, les résultats que nous avons présentés dans ce chapitre indiquent que cette approche permet des résultats similaires aux autres, en particulier lorsque la recherche de patrons est régie par des données additionnelles. Nous aurons donc recours à des informations complémentaires d'ordre sémantique, ainsi qu'à la définition de règles pour les prendre en compte.

Par contre, la méthode des patrons présente certaines limites, qui expliquent pourquoi elle n'est pas plus utilisée.

L'un de ces problèmes est la stabilité des patrons obtenus. Le terme de stabilité désigne ici le nombre de patrons différents pouvant être dérivés à partir d'un phénomène donné. Nous avons vu dans le chapitre 4 que, dans la plupart des études sur les disfluences, un très grand nombre de patrons différents est obtenu pour représenter les disfluences, ou même un seul type de disfluences, telles que les auto-corrections par exemple. Du fait de cette multiplicité de patrons à traiter, une prise en compte extensive impliquerait un algorithme lourd à mettre en œuvre. Mais comme nous l'avons également montré dans le chapitre 4, ce problème est moins prégnant en ce qui concerne les disfluences qui apparaissent dans le dialogue de Contrôle Aérien.

Un autre problème est celui du taux de couverture des patrons, autrement dit le nombre de phénomènes pouvant être modélisés par des patrons. Les productions orales spontanées « tout venant » comportent en effet un certain nombre de manifestations de disfluences

trop complexes pour être représentées par des patrons tels que nous les avons définis dans ce chapitre. Il peut s'agir, par exemple, d'auto-corrections comportant plusieurs niveaux d'imbrication, ou composées d'un nombre important d'éléments. Ce type est alors ignoré par tout traitement basé sur des patrons. Cependant, en nous référant au chapitre 4, p. nous voyons à nouveau que ce n'est pas le cas dans le type de dialogue qui nous intéresse ici. En effet, nous y observons que, pour les auto-corrections, 3 patrons couvrent à eux seuls 67% des occurrences. Et seulement 6% des patrons d'auto-corrections présentent une variabilité trop grande pour être représentés par des patrons stables. Pour les répétitions, le taux de couverture est encore plus important, puisque 4 patrons permettent de couvrir 100% des occurrences relevées.

En raison de ces limites, on peut s'interroger sur la portabilité de la technique des patrons à d'autres types de dialogues. Ce dernier point ne nous semble pourtant pas poser problème. En effet, comme nous l'avons montré plus haut, cette technique a été appliquée avec succès sur des dialogues « quotidiens » ; et pour les raisons que nous venons de citer, elle nous paraît particulièrement appropriée à des dialogues contraints tels que ceux relatifs au Contrôle Aérien.

Ces points étant établis, il reste à déterminer plus précisément le *modus operandi* que nous envisageons. Un point préliminaire est de distinguer des traitements différents selon la nature des disfluences considérées. Nous partons de la constatation que les disfluences peuvent ou non affecter la structure de l'énoncé dans lequel elles apparaissent. Ainsi, ce que nous avons appelé les fragments de mots, les hésitations, ou les allongements, n'affectent pas la structure de l'énoncé<sup>85</sup>. Au contraire, chaque occurrence de répétition, amorce, ou auto-correction induit une modification de l'énoncé par rapport à la planification initiale par le locuteur.

Il découle de cette observation qu'on ne peut envisager une identification générique pour tout type de disfluence.

Pour la première catégorie (fragments de mots, hésitations, allongements), il s'agira uniquement de détecter l'occurrence elle-même, sans aller plus loin. Nous considérons donc que seules l'identification de leurs occurrences est nécessaire et suffisante pour un traitement automatique (quelle que soit l'utilisation faite par la suite de leur identification).

En revanche, les répétitions, amorces, ou auto-corrections présentent une complexité qui nécessite un traitement plus particulier. Elles affectent en effet un empan plus important en termes de taille que la seule position dans laquelle elles apparaissent. Ainsi, il ne s'agit pas

---

<sup>85</sup> On pourrait objecter ici que l'apparition de ce type de disfluences est souvent concomitante avec des perturbations plus larges du contexte où elles apparaissent. Cependant, ces relations sont-elles de cause ou de conséquence ? La question reste à trancher... En attendant, nous adoptons comme hypothèse de travail l'indépendance de causalité des phénomènes entre eux.

seulement d'identifier la disfluence elle-même, mais également la zone dans laquelle celle-ci se produit, et les éléments concernés dans l'énoncé. Plus précisément, il s'agit d'identifier :

- La borne ouvrante de la disfluence : quel est l'élément lexical qui marque le début de la zone disfluente ;
- La borne fermante de la disfluence : quel est l'élément lexical ou pas qui marque la fin de la zone disfluente, à partir de laquelle le reste de l'énoncé est produit normalement ;
- La zone d'édition : elle est à entendre au sens que nous avons défini précédemment.

En plus de ce cadre commun, chaque type de disfluence présente des spécificités qui impliquent de rechercher des éléments particuliers :

- **Répétitions :**
  - Le répétable, autrement dit l'élément faisant l'objet de la répétition disfluente ;
  - Le ou les répétés.
- **Amorces non correctives :**
  - L'amorce ;
  - Le mot correspondant à l'amorce, prononcé de manière complète.
- **Auto-corrections et amorces correctives :**
  - Le ou les éléments corrigés ;
  - La ou les corrections faites par le locuteur : le ou les éléments apportant une correction sont différents de le ou les éléments corrigés : dans le cas contraire ou aurait affaire à une répétition.

La synthèse des différents types de recherche à effectuer sur les disfluences est présentée dans le Tableau 28 ci-dessous.

**Tableau 28** : Comportements de recherche à adopter selon le type de disfluences

	Hésitations, fragments de mots, allongements	Répétition	Amorce non corrective	Auto-correction (avec amorce)	Auto-correction (sans amorce)
Recherches à effectuer	- La disfluence elle-même	- L'élément <sup>86</sup> faisant l'objet de la répétition (le répétable)  - Le ou les éléments répétés	- La ou les amorces  - Le ou les éléments complets	- Borne ouvrante  - Borne fermante  - Point d'édition  - L'élément corrigé ( <i>i.e.</i> l'amorce)  - La correction	- Borne ouvrante  - Borne fermante  - Point d'édition  - L'élément corrigé  - La correction

Du fait des contraintes de la phraséologie, nous n'avons pas à mettre en œuvre de dispositifs spécifiques pour le traitement de phénomènes non disfluents ambigus, fréquents dans le langage « tout venant ». Nous pensons notamment au problème des répétitions d'emphase, que nous évoquons dans le chapitre 1 (sections 3.2.4.4 et 3.2.5.2), ou encore des énumérations (cf. Johnson *et al.* (2004a) ou Guénou (2005, p. 331)).

### 3.2.2. Méthode de recherche

Une fois que l'on a défini les différents éléments à rechercher, il reste à déterminer comment mettre en œuvre leur recherche.

Nous choisissons d'adopter une méthode de recherche locale. Nous employons ce terme par opposition à une recherche globale, qui porterait sur l'ensemble de la phrase. Ce choix est motivé par notre analyse des patrons de formation des disfluences. Aucun ne touche l'ensemble d'un énoncé ; on a de plus constaté (chapitre 4) que la taille des patrons est restreinte. L'analyse locale est donc plus appropriée.

En ce qui concerne le moteur de recherche des patrons à proprement parler, nous sommes là encore guidés par le souci de la méthode la plus simple et adaptée au modèle de tâche du contrôle aérien. La réponse est une fois encore fournie dans les articles de Bear *et al.* (1992), Heeman *et al.* (1994). Comme ces auteurs, nous nous proposons de reconnaître les patrons à l'aide d'une série de règles dont le rôle est de compléter et cadrer l'utilisation des patrons et éviter les « faux positifs ». L'utilisation de règles au sein d'algorithmes est d'ailleurs employée dans d'autres approches de traitement des disfluences, telles que celle de Hindle (1983), Kurdi (2003) et Goulian (2002), même si pour ces deux derniers auteurs, les règles jouent surtout le rôle de compléments pour traiter des cas particuliers.

<sup>86</sup> Nous employons le singulier car, comme nous l'avons montré dans le chapitre 3, notre corpus ne comprend aucune occurrence de répétition avec plus d'un élément répétable.

Ainsi, l'approche que nous proposons peut se résumer en quatre règles principales :

1. Repérer une disfluence et/ou une zone susceptible d'en contenir une ;
2. Mener une série de tests pour déterminer si il s'agit bien d'une disfluence ;
3. Si le phénomène détecté est bien une disfluence, identifier de quel type il s'agit ;
4. Enfin, selon le type identifié, faire les recherches appropriées, synthétisées dans le Tableau 28.

Chacune des actions décrites par ces règles est elle-même décomposable en un sous-ensemble de règles. Leur description détaillée est assimilable à un algorithme. Conformément aux objectifs que nous avons définis pour ce chapitre, nous ne rentrons pas plus dans ces détails pour l'instant.

#### **4. Conclusion**

Dans ce chapitre, nous avons exploré les principales approches employées dans la littérature pour la détection automatique des disfluences.

Nous avons ensuite mis en perspective cette connaissance par rapport aux spécificités de notre corpus. Nous avons ainsi vu qu'un traitement par règles et reconnaissance de patrons est le plus pertinent compte tenu de ces contraintes.

Nous avons enfin présenté les principes que nous proposons de mettre en œuvre pour la détection automatique des disfluences dans les dialogues de contrôle aérien. Ils diffèrent selon la nature de la disfluence à traiter. Pour les phénomènes complexes, nous proposons une analyse locale, avec de surcroît un contrôle sémantique, indispensable pour une plus grande robustesse.

Cependant, nous sommes restés à un niveau assez théorique, puisque nous n'avons pas envisagé les modalités exactes de réalisation. Il nous reste à présent à envisager l'implémentation de ces principes. Cela implique leur traduction dans des algorithmes, qui seront eux-mêmes intégrés dans le cadre d'un système informatique. Nous présentons l'implémentation que nous avons effectuée dans le chapitre suivant.



## **Chapitre 6**

### **Implémentation**

#### **1. Introduction**

Dans le chapitre précédent, nous avons envisagé d'un point de vue théorique la méthodologie que nous proposons pour détecter automatiquement les disfluences dans le contexte du dialogue de contrôle aérien. Cependant, ce n'est qu'une première étape dans un processus de développement informatique. Il est en effet nécessaire de passer de la description des principes adoptés à leur réalisation concrète dans un système informatique, autrement dit leur implémentation.

Ceci est l'objectif du présent chapitre. Il est consacré à la description précise de l'implémentation du modèle et des règles que nous proposons. Nous présentons également le système existant dans lequel l'implémentation est réalisée. En effet, un dispositif de traitement des disfluences doit se placer dans un système complet de reconnaissance et de compréhension de la parole pour disposer des données à traiter.

Dans un premier temps, nous définissons les spécifications de l'implémentation que nous proposons. Nous décrivons ensuite l'implémentation. Dans la deuxième section, nous présentons le cadre de celle-ci, tandis que dans la troisième nous nous penchons sur ses détails.

#### **2. Spécifications de l'implémentation**

On peut fixer un nombre minimum de spécifications définissant ce que doit pouvoir accomplir l'implémentation que nous proposons :

1. Correspondre à la modélisation ainsi qu'aux différents phénomènes présentés dans le chapitre précédent ;
2. Etre utilisable aussi bien pour prendre en compte la parole orale spontanée que des transcriptions au format électronique (que celles-ci soient produites en temps réel, par exemple par un logiciel de reconnaissance du signal, ou viennent d'un fichier existant). D'un point de vue technique, être aussi portable que possible quels que soient la plateforme matérielle ou le système d'exploitation utilisés ;

3. Etre adaptée au contexte particulier dans lequel nous situons ce travail : le contrôle de trafic aérien ;
4. Présenter un temps de réactivité suffisamment bas pour permettre une application temps réel.

### **3. Présentation du cadre d'implémentation**

#### **3.1. Historique et objectifs**

Le système dans lequel nous allons insérons notre modèle a été développé dans le cadre du projet VOICE. Nous avons déjà présenté ce projet dans le chapitre 3. Comme nous l'avons souligné alors, l'un des buts est de permettre, lors de l'entraînement des contrôleurs aériens, de décharger les pseudos pilotes de certains des vols virtuels qu'ils doivent simuler, au profit d'un système automatique.

Plus précisément, le but de ce système est de gérer les entrées vocales pouvant être produites par les pseudo-pilotes, en extraire le sens, et l'interpréter de manière à ce que les ordres reçus soient exécutés. Comme nous le montrerons, il a déjà fonctionné dans des conditions de simulation du contrôle aérien, même si un certain nombre de possibilités restent à prendre en compte, notamment la gestion des disfluences et autres problèmes dans le dialogue !

Pour arriver à ce résultat, le système dispose d'une architecture particulière que nous allons présenter en détails dans la section qui suit.

#### **3.2. Architecture**

Le système dans son ensemble dispose d'une architecture modulaire, puisqu'il fait intervenir différents agents communiquant entre eux. Ces agents sont des petits programmes dédiés chacun à une sous-tâche particulière requise pour l'exécution correcte de la compréhension et de l'exécution des ordres correspondants. Avant de présenter plus précisément le rôle de chacun d'entre eux, nous décrirons l'environnement général dans lequel ils sont exécutés, pour ensuite nous pencher sur la gestion de l'oral par VOICE.

##### *3.2.1. Gestion de la communication inter-agents*

Pour permettre la communication entre les agents, il a été choisi de recourir à un *middleware* (ou *intergiciel*, ou *bus logiciel*, en français) nommé *IVY*. Rappelons qu'un *middleware* est un logiciel permettant la communication entre différentes applications. Le problème à résoudre est que les applications considérées peuvent ne pas être développées dans le même langage de programmation, ne pas utiliser le même format de données en entrée/sortie, ne pas être exécutées/exécutables sur le même système d'exploitation, etc. Il

est donc nécessaire de leur fournir une sorte de *lingua franca*<sup>87</sup> pour communiquer entre elles, ce que permet le middleware. Pour cela, il fournit un certain nombre de fonctionnalités, notamment au niveau de la couche réseau.

IVY est un *middleware* développé au CENA dans l'équipe PII. La majorité des informations que nous allons présenter ici peut être retrouvée dans Chatty *et al.* (2002) et Buisson *et al.* (2002), ainsi que sur Internet<sup>88</sup>. IVY est développé au dessus des couches TCP et UDP de la pile de protocoles TCP/IP et fournit une abstraction de type « bus logiciel ». En effet, chaque agent, une fois « connecté au bus IVY » émet des messages textuels sans se soucier des destinataires éventuels et reçoit les messages d'autres agents pour lesquels il est abonné. La première caractéristique d'IVY est d'avoir un fonctionnement totalement distribué, sans serveur central. Cela permet une certaine robustesse puisque chaque agent possède un cycle de vie autonome vis-à-vis des autres agents présents dans l'environnement. D'autre part, IVY définit un certain nombre de fonctionnalités, permettant aux applications de se connecter et déconnecter du bus et de communiquer. L'abonnement aux messages est géré sous la forme de *souscription à base d'expressions régulières (regex)* du processus ou de l'application à IVY.

Une fois connectés au bus, les logiciels pourront aussi bien recevoir des données circulant sur IVY qu'en émettre. Nous présentons les différentes applications connectées dans le cadre de VOICE en 3.2.3.

#### 3.2.2. Gestion de la composante orale

Il va de soi que dans le cadre que nous avons décrit pour VOICE, il est nécessaire que la composante orale soit prise en compte. Pour cela, il faut en premier lieu disposer d'un logiciel de reconnaissance vocale. C'est *Dragon Naturally Speaking* 7<sup>89</sup> (DNS) qui a été choisi. Le taux de reconnaissance annoncé est de 98%, pour une dictée vocale effectuée par un utilisateur ayant effectué plusieurs phases d'apprentissage pour le modèle de langage. Evidemment, ce taux est moindre dans la situation qui nous intéresse : le contrôle aérien, y compris le pseudo-pilotage, s'effectue dans un contexte beaucoup moins contrôlé que celui d'une dictée vocale (environnement bruité, contexte de dialogue ...). De plus, il est impossible de faire passer une phase d'apprentissage à chaque utilisateur de VOICE : cela serait très couteux en temps. De ce fait, un modèle générique est utilisé, ce qui est par conséquent moins performant<sup>90</sup>.

---

<sup>87</sup> Le langage commun utilisé pendant des siècles par les marins pour communiquer indépendamment de la langue natale ou des origines culturelles.

<sup>88</sup> Voir notamment <http://www.tls.cena.fr/products/ivy/>.

<sup>89</sup> <http://www.nuance.fr/naturallyspeaking>. En 2007, la dernière version est la 9 ; mais pour des raisons techniques (notamment incompatibilité avec *SAPI 4*), nous sommes contraints d'utiliser la version 7.

<sup>90</sup> Il n'y a pas à notre connaissance d'études indiquant le taux de reconnaissance de DNS dans le cas de l'utilisation d'un tel modèle.

D'autre part, il a été nécessaire de mettre en œuvre un module permettant de gérer l'interface entre les sorties de DNS et les autres agents utilisés dans VOICE. On rappelle en effet que tout moteur de reconnaissance vocale reçoit en entrée un signal audio, et produit en sortie une ou plusieurs propositions correspondant à l'interprétation du signal. Le rôle de ce module est donc de recevoir les différentes solutions proposées par DNS, et de les envoyer à IVY pour traitement. Son développement a été assuré en utilisant *SAPI 4* (*Speech Application Programming Interface*) de Microsoft<sup>91</sup>. Comme son nom l'indique, il s'agit d'une bibliothèque permettant de gérer les entrées/sorties vocales sur un PC.

Maintenant que nous avons présenté l'environnement de fonctionnement de VOICE, il nous reste à en décrire précisément l'architecture et le mode de fonctionnement.

### 3.2.3. Description des principaux agents

Il y a 4 principaux agents, chacun correspondant à un aspect particulier du traitement à effectuer :

- Le module *Sra* : nous avons déjà présenté son rôle ci-dessus ; on rajoutera ici que ce module, pour décoder le signal reçu, il recourt à une grammaire décrivant des règles de réécriture. Cette grammaire est dynamique : elle est modifiée à chaque entrée ou sortie d'un avion dans le ou les secteurs contrôlés à partir d'un fichier source définissant les règles de base. Ces règles concernent notamment les paramètres de vol courants. Le format des règles et leur utilisation détaillée étant évidemment d'une grande importance pour l'implémentation de notre modèle, nous les exposerons dans la section 3, qui lui est consacrée ;
- *Autogram* : cet agent sert à générer automatiquement, sous la forme d'une grammaire, l'ensemble des prononciations possibles pour un indicatif donné. Rappelons que chaque indicatif correspondant à un vol, constitué d'une séquence de lettres et chiffres, peut être prononcé à l'oral de différentes manières. *Autogram* permet de prendre en compte ces différentes possibilités de prononciation. Les résultats produits sont traités par le module *Sra*. *Autogram* a été développé par L. Dourmap et est présenté dans son mémoire de DEA (Dourmap, 2002, *op. cit.*) ;
- *Voice\_decision* : on peut dire que c'est l'agent « pivot » de VOICE. En effet, c'est lui qui reçoit de *Sra* les différentes solutions du moteur de reconnaissance, et choisit la solution qui est la plus pertinente (c'est-à-dire la plus proche possible de ce qui a été effectivement dit par le locuteur). Le choix est fait en fonction de divers critères, autant techniques (position de l'énoncé dans le classement des solutions de *Sra*) que contextuels (notamment historique du dialogue et paramètres des

---

<sup>91</sup> <http://www.microsoft.com/speech/speech2007/default.msp>. La version actuelle est la 5.x. Cependant, à l'époque où VOICE a été conçu, c'est la version 4.x qui était en vigueur. Cela a conditionné le choix de la version utilisée de Nuance, comme nous l'expliquons ci-dessus.

différents vols simulés à l'instant du calcul). Une fois que le choix a été fait, la décision est envoyée sur IVY ;

- *History Orders* : permet la sauvegarde des différentes sorties envoyées sur IVY par *Voice\_decision*.

D'autres composants logiciels participent également à diverses sous-tâches exécutées pour VOICE. Deux d'entre eux sont notables vu leur importance dans l'utilisation du dispositif. Ils permettent en effet, respectivement, de prendre en compte les connaissances contextuelles, et d'afficher les vols virtuels.

Le module permettant « d'injecter » des connaissances contextuelles sur les vols simulés, notamment les identifiants, les positions, le temps, etc. sert également à gérer de manière dynamique l'évolution de tous les vols, ainsi que, le cas échéant, les modifications provoquées par les ordres donnés par le contrôleur<sup>92</sup>.

Plus précisément, l'utilisation de la connaissance du contexte dans VOICE peut-être schématisée comme dans la Figure 20 ci-dessous (Truillet *et al.* 2002, p. 181).

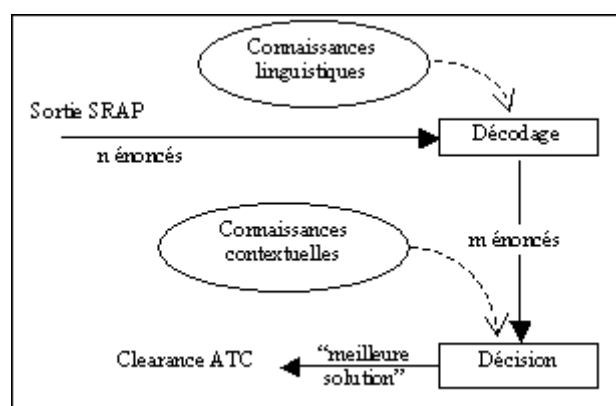


Figure 20 : Gestion des connaissances linguistiques et contextuelles dans VOICE.

Comme on le voit sur cette figure, il faut bien distinguer entre les connaissances d'ordre linguistique et les connaissances contextuelles au sens propre. Les premières consistent principalement en l'utilisation de *Context Free Grammar* (CFG). Elles permettent par exemple de développer les différentes possibilités de reformulation des indicatifs. Elles interviennent en sortie du système de reconnaissance, au niveau de *Sra*. Comme l'implémentation que nous proposons fait beaucoup appel à ces règles, il est nécessaire de les décrire ici plus en détail.

Une définition formelle<sup>93</sup> pour commencer : une grammaire CFG est un *tuple* G tel que

<sup>92</sup> Il est à noter que les ordres, lorsqu'ils sont compris, sont systématiquement appliqués, sans qu'il y ait gestion des risques éventuels (mais virtuels) que l'ordre pourrait entraîner s'il est faux !

<sup>93</sup> Le créateur est N. Chomsky, qui en fait la première description dans son article Chomsky (1956), disponible à l'url : <http://www.chomsky.info/articles/195609---.pdf>.

$$G = \{ V, \Sigma, R, S \}$$

où :

**V** : Ensemble des éléments non-terminaux

**S** : Ensemble des éléments terminaux

**R** : Relation de V vers V ou  $\Sigma$ .

**S** : élément de départ pour l'analyse.

Concrètement, ce *tuple* se manifeste par une série de règles dites de réécriture, par lesquelles un composant initial est décomposé en sous-composants, et ainsi de suite selon un procédé qui peut généralement être récursif<sup>94</sup>. L'une des principales applications de ce type de grammaire est la formalisation et le traitement du langage naturel<sup>95</sup>. L'inconvénient est par contre un pouvoir expressif parfois limité pour rendre compte de certains phénomènes. Comme nous le verrons plus bas, c'est une des limitations que nous avons rencontré dans le traitement des disfluences.

Elle est appelée « *Context Free* » car l'apparition d'un mot n'est pas subordonnée au contexte, c'est-à-dire aux mots voisins. On se situe ainsi par opposition à d'autres méthodes, comme les *Modèles de Markov Cachés*, pour lesquelles un état donné dépend des *N* précédents états parcourus. L'un des principaux avantages des CFG est une grande souplesse d'utilisation. D'autre part, contrairement aux méthodes statistiques, les CFG ne nécessitent pas d'être au préalable entraînées sur un nombre important de données<sup>96</sup>. Nous revenons sur les détails pratiques de l'utilisation que nous avons faite des CFG dans une section ultérieure.

Les connaissances contextuelles, quant à elles, relèvent notamment des ordres déjà donnés (l'historique du dialogue), des paramètres de chaque vol à un moment donné (position, vitesse, niveau), et d'autres types d'information : limites minimum et maximum des vols, mode de formation des indicatifs, etc. Le but est d'aider au maximum à la décision parmi les différentes solutions en sortie de *Sra*, en éliminant celles ne correspondant pas aux données réelles de la situation. Par exemple celles comportant un indicatif inexistant ou encore un niveau de vol interdit.

Le deuxième module que nous évoquions sert à gérer la transmission de certaines de ces informations à une interface permettant de les consulter de manière plus conviviale, que ce soit sous forme graphique (cf. ci-dessous) ou textuelle. Il permet également, si besoin est, de se substituer à la modalité vocale, en envoyant sur IVY des messages qui seront traités

---

<sup>94</sup> Sauf si R ne porte que, et uniquement que, sur S.

<sup>95</sup> C'est d'ailleurs le but initial de N. Chomsky. Mais signalons toutefois que leur champ d'utilisation est très vaste, et concerne notamment les mathématiques ou la bioinformatique.

<sup>96</sup> Cf. notamment Cornuéjols *et al.* (2003), chapitres 2 et 5.

### 3. Présentation du cadre d'implémentation

comme s'ils provenaient de *Sra*. La seule contrainte dans ce cas est de respecter le format de sortie de *Sra* (dont nous donnons un exemple plus bas).

L'architecture globale du système est résumée dans la Figure 21 ci-dessous, empruntée à Truillet (*et al.* 2005, p. 74).

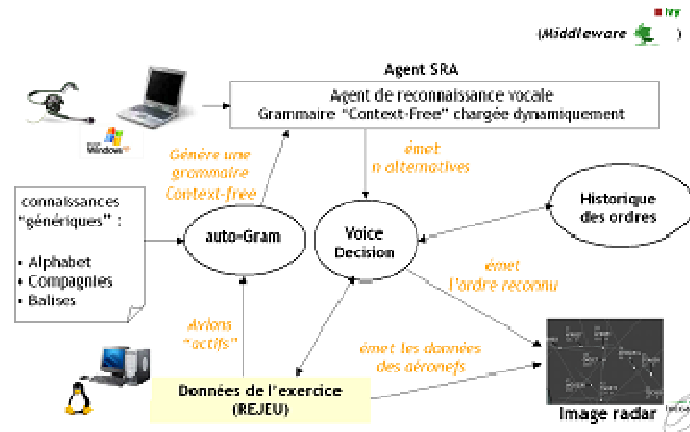


Figure 21 : Architecture de VOICE

D'un point de vue matériel, l'équipement minimal pour faire fonctionner VOICE est un ordinateur doté d'un microphone, suffisamment puissant pour faire tourner les différents logiciels à l'œuvre (y compris le moteur de reconnaissance). Il sera éventuellement complété d'une représentation visuelle, semblable aux images radars utilisées dans le contrôle aérien, pour permettre à l'utilisateur une meilleure visualisation des différents vols simulés, et de l'impact des ordres donnés.

Un exemple de fonctionnement est fourni dans (Truillet *et al.* 2005). Soit l'énoncé

Correction Jersey trente trente-neuf, montez niveau 300

Son traitement par VOICE suivra les étapes suivantes :

- Après analyse par DNS du signal correspondant, le module *Sra* va proposer les deux solutions de décodage les plus probables (nous indiquons la signification de chaque élément des deux solutions) :
 

sra	Parsed=	Correction:CORRECT	callsign=JEA3039	Niveau:FL_CLB=300	NP=1 Num_A=1
		« Correction »	« Jersey »	« Montez niveau 300 »	Numéro de la solution
sra	Parsed=	Correction:CORRECT	callsign=JEA3039	Niveau:FL_CLB=100	NP=1 Num_A=2
		« Correction »	« Jersey »	« Montez niveau 100 »	Numéro de la solution
- Ces solutions sont envoyées sur IVY ;
- Voice\_decision*, en écoute sur IVY, va recevoir ces deux solutions et les traiter de manière à choisir la plus pertinente compte tenu du contexte. Pour ce

faire, le module va procéder à plusieurs vérifications concernant les différents paramètres du vol considéré : niveau courant, vitesse, etc. Dans notre exemple, il s'avère que le vol se trouve, au moment où l'ordre est donné, au niveau 195. La deuxième solution ne peut donc être retenue, puisqu'elle implique une contradiction entre le fait que l'ordre est de monter, et que le niveau proposé est inférieur au niveau courant. De ce fait, seule la première solution est viable. La décision est envoyée sur IVY ;

- 4) La situation simulée évolue en fonction de l'ordre choisi : le vol *JEA3039* monte du niveau 195 au niveau 300. Les modifications correspondantes dans les représentations visuelles et/ou textuelles sont actualisées. L'ordre est également sauvegardé dans l'historique du dialogue pour une éventuelle utilisation ultérieure ; par exemple pour que *Voice\_decision* puisse vérifier que des solutions proposées par *Sra* ne sont pas incohérentes avec un ordre donné précédemment.

Ce système a donné lieu à quelques évaluations pour les parties déjà implémentées. Pour que le lecteur puisse avoir une idée aussi bonne que possible des capacités du système, nous présentons brièvement les résultats déjà obtenus dans la prochaine section.

### **3.3. Capacités et limites de VOICE**

Pour l'instant, deux principales évaluations ont été accomplies, et concernent le travail de L. Dourmap sur la génération des différentes prononciations pour un indicatif donné. Elles sont décrites dans Dourmap (2002, p. 37-56) et Truillet *et al.* (2005, p. 71-73). Le but était principalement d'évaluer les intérêts et inconvénients de deux grammaires différentes, ainsi que, d'une manière générale, le comportement du système dans son ensemble.

La première de ces deux grammaires est dite *générique*. Elle rend compte de toutes les prononciations possibles de n'importe quel indicatif. La seconde grammaire est appelée *développée*. Contrairement à la précédente, elle utilise des connaissances contextuelles, à savoir la liste des vols simulés, et leurs indicatifs, pour limiter le nombre de prononciations possibles.

Deux principaux paramètres ont été employés pour les évaluations. Ce sont :

- Le taux d'erreur : plusieurs mesures de taux d'erreurs ont été menées. Elles varient selon divers paramètres, notamment la ou les réponses considérées. Dans tous les cas, la grammaire développée présente des taux d'erreurs moins importants que la grammaire générique ;
- Le temps mis par le système pour traiter l'information : là encore, la grammaire développée est plus intéressante, car plus rapide, que la grammaire générique. Lorsqu'elle est utilisée, le temps est suffisamment court pour pouvoir être



intéressant y compris dans le cadre d'une application en temps réel (inférieur à 700 ms au pire cas).

En ce qui concerne le modèle de langage, le système VOICE est relativement limité. En effet, d'une part il ne concerne que les ordres dits simples ; d'autre part, seulement quatre paramètres du contrôle aérien sont pris en compte : le niveau, la direction, la vitesse, et la localisation. Cependant, il constitue malgré tout une base appréciable pour notre travail. En effet, les limites que nous venons de dresser ne sont que très relatives : les quatre paramètres pris en compte sont très utilisés dans les dialogues de contrôle aérien, et couvrent donc une grande majorité de ce qui peut être exprimé par la phraséologie. La même remarque peut être faite pour la proportion d'ordres simples par rapport aux ordres complexes. Enfin, VOICE permet déjà de prendre en compte certains phénomènes qui nous intéressent dans le cadre de ce travail : certains marqueurs de politesse pour l'ouverture du discours, tels que « bonjour », ainsi que de l'auto-correction par le contrôleur. Précisons cependant que pour ce dernier cas, ne sont concernés que les cas où le mot « correction » est prononcé antérieurement à la correction elle-même (l'élément corrigeant une erreur faite précédemment). Le nombre de cas potentiellement concernés est donc assez peu élevé (se reporter aux statistiques du chapitre 4). Mais l'approche est déjà intéressante et nous évite d'avoir à nous en occuper. L'une des lacunes de VOICE est justement l'impossibilité de traiter « les phénomènes extra-linguistiques, d'auto-correction et des déviations par rapport à la phraséologie officielle observées in situ », comme cela est noté en conclusion de Truillet *et al.* (2005, p. 74).

Voyons à présent comment nous proposons d'implémenter les principes décrits au chapitre précédent.

#### 4. Implémentation de notre modèle

Les modifications que nous apportons au système VOICE sont de deux ordres. D'une part au niveau de la grammaire utilisée, et d'autre part par la création d'un agent supplémentaire. Cette distinction reflète également la distinction, que nous avons établie dans le chapitre précédent, entre d'un côté les disfluences ne modifiant pas la structure de l'énoncé, et celles le faisant. Les premières sont traitées au niveau de la grammaire, les secondes au niveau du nouvel agent, que nous avons nommé *module\_disfluences*.

Dans les sous-sections qui suivent, nous décrivons respectivement ces deux niveaux de modification.

##### 4.1. Modifications au niveau de la grammaire

Les modifications que nous avons apportées à la grammaire sont de deux types :

- Prise en compte des hésitations et autres disfluences de ce type. Pour cela, nous avons rajouté des règles correspondant à la production de ce type de phénomènes. Pour deux raisons techniques, nous n'avons pas pu prendre en compte la catégorie des pauses vides. En effet, il est pour l'instant impossible de distinguer les pauses vides telles que nous les avons définies et une absence de production liée à la fin de l'énoncé : par défaut, une pause silencieuse trop longue correspond à la fin du traitement par VOICE. D'autre part, il n'existe aucun symbole permettant de rendre compte d'un silence, contrairement à d'autres phénomènes. De même, le fonctionnement du moteur de reconnaissance empêche la prise en compte des allongements. En revanche, nous avons mis en place la prise en compte des pauses remplies, des négations, et de certains mots utilisés pour combler des vides (« bon », « alors », « etc. ») ;
- Introduction de la dimension sémantique. Comme nous l'avons expliqué plus haut (section 4.2.1), il est nécessaire de rendre compte du sens véhiculé par les différents éléments traités. D'une part, parce que dans l'architecture d'IVY, la grammaire est le module permettant d'ajouter du sens aux éléments lexicaux traités de la façon la plus économique. Et d'autre part, parce que les règles de formation de la grammaire de SAPI 4 permettent de procéder relativement facilement à cette introduction de la sémantique.

## **4.2. Conception et réalisation de l'agent**

### **4.2.1. Principe général de l'algorithme**

Pour décrire d'un point de vue général le traitement que nous proposons, nous avons utilisé une description en quatre points dans le chapitre 4. Nous la reprenons ci-dessous, en détaillant chaque point.

#### **1. Repérer une disfluence et/ou une zone susceptible d'en contenir une**

Comme nous le verrons plus bas, le repérage des disfluences « simples » telles que les hésitations se fait au niveau de la grammaire. L'agent *module\_disfluences* intervient pour le repérage des disfluences complexes.

Pour comprendre son principe de base, revenons sur l'analyse des différents patrons de ces disfluences. Dans tous les cas, nous constatons une symétrie des composants situés de part et d'autre du point d'édition. Cette observation va nous servir de point d'achoppement pour la mise en œuvre de l'implémentation : repérer une disfluence complexe va d'abord revenir à chercher dans l'énoncé une symétrie entre deux éléments lexicaux.

Comment parcourir l'énoncé ? Nous proposons une méthode d'analyse locale à l'aide d'une fenêtre coulissante. La fenêtre est d'une taille de 10 éléments lexicaux. Cette taille

peut sembler importante, compte tenu de la taille moyenne des énoncés. Mais elle est motivée par certains patrons que nous avons trouvés, d'une taille totale supérieure à 10 éléments. Certes, ces patrons sont minoritaires. Mais étant donné qu'il n'y a aucun problème à utiliser une fenêtre de cette taille sur un énoncé de moins de 10 éléments, nous préférons conserver ce paramètre de taille. On remarquera que cet algorithme est basé sur l'observation de comportements et de déviations locales, puisque l'empan considéré est inférieur à 10 « mots ». On pourrait objecter que les disfluences peuvent perturber un empan beaucoup plus large de l'énoncé, surtout lorsqu'elles se produisent dans le cadre de situations « normales ». Mais justement, d'une part, les disfluences que nous observons se situent dans un contexte très contrôlé ; d'ailleurs la taille moyenne d'un énoncé est d'un peu plus de 10 mots. D'autre part, comme indiqué dans le chapitre 1, il est avéré que les troubles induits par auto-corrections sont dans la plupart des cas, restreints à un empan limité, sauf cas pathologiques.

La fenêtre se déplace de gauche à droite de l'énoncé, du début à la fin de celui-ci. Le principe est de détecter, dans l'intervalle couvert par la fenêtre, l'occurrence d'une répétition. Plus précisément, on s'intéresse à deux types de répétitions : soit répétition d'un élément lexical (par exemple « montez montez »), soit répétition de deux éléments lexicaux différents, mais appartenant à une même catégorie sémantico-fonctionnelle (par exemple, « montez descendez », où les deux verbes sont différents, mais expriment tous les deux un ordre relatif au déplacement du vol par rapport à un niveau. Nous rejoignons ici l'idée d'un « contrôle sémantique » pour augmenter l'efficacité de la recherche de patrons, utilisée notamment dans Bear *et al.* (1992, p. 59-60) et Bove *et al.* (2007, p. 107-110)<sup>97</sup>. Comme on s'en doute, le contrôle sémantique repose sur la mise en correspondance de chaque élément lexical du lexique à sa catégorie sémantico-fonctionnelle. Cette mise en correspondance est réalisée au niveau de la grammaire *CFG*.

Si la présence d'une répétition est trouvée dans une fenêtre, on procède aux traitements décrits dans les points suivants de la procédure, avant de poursuivre le déplacement. Si aucune répétition n'est détectée, la fenêtre se déplace d'un élément vers la droite. Pour gérer le déplacement de la fenêtre, nous avons le choix entre deux principales solutions<sup>98</sup>, une fois que la limite droite de la fenêtre a été atteinte sans trouver de répétitions :

- La fenêtre se déplace d'une seule position vers l'avant. Autrement dit, toutes les positions qu'elle contient sont incrémentées de +1 ;

---

<sup>97</sup> Bien que dans ce dernier cas, le contrôle sémantique concerne une autre approche que la recherche de patrons.

<sup>98</sup> Nous utilisons le terme de « principales » car d'autres solutions seraient envisageables, en fonction de la taille du pas de déplacement choisi : 2, 3, etc. Nous nous limitons à l'exposition de ces deux solutions car elles sont les plus pertinentes dans le contexte qui nous intéresse.

- La fenêtre se déplace à la suite de la dernière position précédente. Autrement dit, l'ensemble des positions qu'elle contient sont incrémentées de +10.

Cette précision est importante car le comportement général du système sera grandement modifié selon le choix effectué. Dans le premier cas, la fenêtre permet de repérer la moindre répétition dès son apparition dans l'énoncé traité. Dans le second, il est possible de rater des occurrences de répétition et donc de correction. En effet, pour reprendre la terminologie de la théorie des ensembles, la première solution permet de calculer l'intersection des deux ensembles formés par les positions respectives des fenêtres au moment  $n1$  et  $n2$ , tandis que la deuxième est juste une juxtaposition. Nous avons donc évidemment choisi la première solution.

## **2. Mener une série de tests pour déterminer si il s'agit bien d'une disfluence ;**

Nous avons montré dans le chapitre 5 qu'un des problèmes rencontrés lors de la recherche automatique des disfluences est la présence de nombreux « faux positifs ». Ce problème est *a priori* encore plus important compte tenu de l'approche que nous avons choisie et du contexte du contrôle aérien. En effet, nous avons défini comme indice de présence d'une disfluence l'occurrence d'une répétition. Or, les répétitions non disfluentes, présentes dans le langage oral spontané non contraint, apparaissent également dans la phraséologie. Dans ce contexte, il est donc nécessaire de s'assurer, selon une méthode spécifique, que l'on n'a pas affaire à un faux positif.

Compte tenu de la phraséologie et de la tâche de Contrôle Aérien, il n'y a que deux cas dans lesquels on peut avoir affaire à une répétition non disfluente<sup>99</sup>, que ce soit au niveau lexical ou sémantico-fonctionnel :

- D'une part, lors de l'expression des indicatifs et celle d'autres indices numériques, tels que les niveaux ou les vitesses. Pour ces deux derniers cas, les répétitions non disfluentes sont causées par la possibilité, offerte par la phraséologie, de prononcer séparément les composants d'un nombre ou d'un indicatif (par exemple, « 1 1 2 » au lieu de « cent douze »)<sup>100</sup>. A chaque fois, il y a répétition des mêmes catégories sémantico-fonctionnelles ; selon les cas, il peut y avoir également répétition au niveau lexical, comme par exemple les deux premiers nombres dans « 1 1 2 » ;
- D'autre part, lors de l'indication d'un chemin composé de plusieurs balises, tel que par exemple « Amboise Poitiers Balon ». Ce cas ne concerne que le niveau

---

<sup>99</sup> Contrairement au français « standard », qui autorise plusieurs types de répétitions non disfluentes (par exemple, stylistique, accentuation, etc., cf. Chapitre 1 *op. cit.*

<sup>100</sup> Cf. chapitre 2, section 3.1.2.

sémantico-fonctionnel. En effet, l'énumération des différentes balises que doit comporter un plan de vol comporte systématiquement une suite de noms de balises, autrement une série de répétitions de catégories sémantico-fonctionnelles.

Hormis ces deux cas particuliers, toute autre occurrence de répétition au niveau lexical et/ou sémantico-fonctionnel est indicatrice de la présence d'une disfluence. Dans ce cas, on passe au point suivant du traitement.

### **3. Si le phénomène détecté est bien une disfluence, identifier de quel type il s'agit ;**

A ce niveau, la recherche de patrons va être plus fine, puisque la configuration du patron reconnu dans la fenêtre indique à quel type de disfluence on a affaire.

On cherche en effet à déterminer si la disfluence est une répétition lexicale. Selon que c'est le cas ou non, on sait que l'on a affaire, respectivement, à une répétition disfluente ou une auto-correction. Dans les deux cas, on peut passer directement au point 4.

### **4. Selon le type identifié, faire les recherches appropriées.**

Dans un premier temps, le programme cherche à déterminer si les éléments disfluents de la zone traitée correspondent à l'un des plus fréquents des patrons simples présentés dans le chapitre 4. Par exemple, dans le cas d'une auto-correction, on cherchera la congruence avec *R1 / R1*, *M1 R1 / M1 R1*, *R1 M1 / R1 M1*, ainsi que les variantes de ces patrons dans lesquelles sont insérés des éditeurs comme les hésitations, ou les négations. Compte tenu de la fréquence observée de ces patrons, la majeure partie des disfluences devrait être détectée à ce niveau.

Il reste cependant le cas des patrons plus complexes, mais également plus rares, pour lesquels la recherche d'une correspondance avec un patron existant n'est pas possible. En effet, le nombre très limité d'occurrences empêche d'en déduire des règles de formation exactes. Par conséquent, il est plus pertinent dans ce cas d'adopter une méthode d'analyse générique, qui soit robuste aux différentes manifestations qui puissent se produire :

- Trouver la position du point d'édition. Pour cela, on lance une analyse par règle de CFG à partir du début de la fenêtre. Lorsque l'analyse se termine avant que l'un des éléments terminaux de la CFG n'ait été atteint, cela indique la présence du point d'édition. En effet, la quasi-totalité des occurrences de disfluences se manifeste par l'interruption en cours de production d'un ordre simple, avant la fin de celui-ci<sup>101</sup>. Comme par exemple, dans « euh F M I **descent level euh euh**

---

<sup>101</sup> Rappelons qu'un ordre complexe est composé de plusieurs ordres simples.

**descendez** niveau 160 ». Le point d'édition est donc concomitant à l'interruption prématurée de l'ordre ;

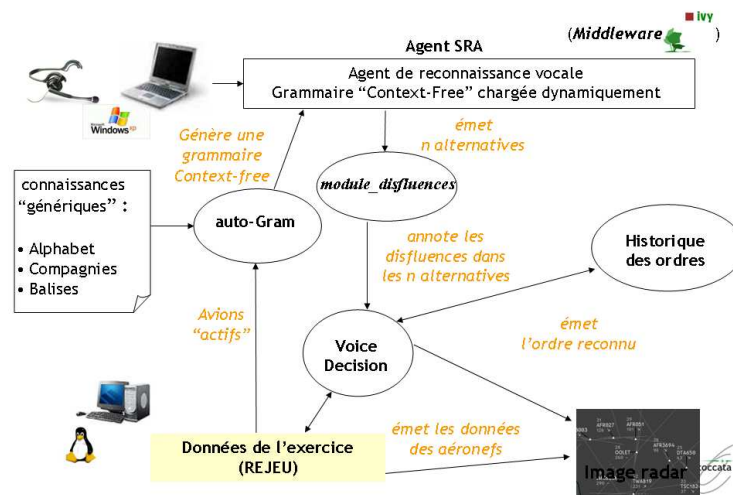
- Le point d'édition trouvé, nous avons à distinguer les éléments répétés entre eux (autrement dit, distinguer les éléments corrigés de leur correction). On fait la comparaison entre les éléments situés à gauche du point d'édition et ceux de droite ; les éléments répétés (d'un point de vue lexical) correspondent ... à des répétitions, tandis que les éléments répétés (au niveau de la classe de mot) correspondent respectivement à une erreur et à sa correction. Une fois identifié ainsi le statut de chaque élément, il est alors possible de leur attribuer automatiquement une étiquette correspondante (par exemple « element\_repete », « repetition », etc.

L'algorithme complet, ainsi qu'un exemple de déroulement, sont présentés respectivement dans les Annexes 1 et 2.

### 4.3. Intégration dans VOICE

#### 4.3.1. Intégration dans l'architecture de VOICE

Deux types d'intégrations différentes dans l'architecture globale de VOICE sont possibles. Elles correspondent à deux prises en compte différentes de la grammaire. La première est illustrée par la Figure 22 :

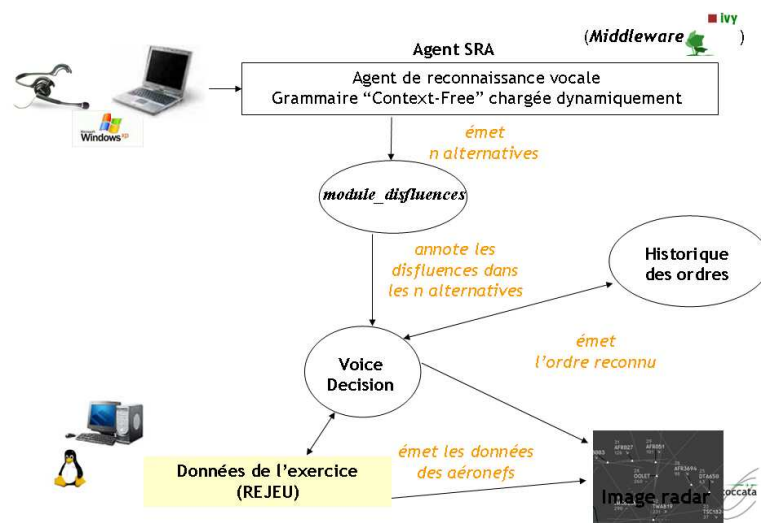


**Figure 22 :** Intégration dans VOICE avec grammaire développée

On voit que *module\_disfluences* est intégré en sortie de *Sra*, et en entrée de *Voice Decision*. Cela signifie que *module\_disfluences* reçoit en entrée les *N* solutions alternatives

envoyées sur *IVY* par *Sra*, chacune correspondant à un énoncé<sup>102</sup>. Chacune de ces solutions est ensuite traitée, avant d'être renvoyée à *Voice Decision*, qui effectue le choix de la plus plausible, selon la procédure décrite en 3.2.3.

La particularité de ce type d'intégration de *module\_disfluences* est qu'elle fait appel à une grammaire développée et au module *auto-Gram*, dont nous avons décrit le rôle et le fonctionnement en 3.2.3. C'est en ceci qu'elle diffère du deuxième mode d'intégration, décrit dans la Figure 23 ci-dessous.



**Figure 23 :** Intégration dans VOICE avec grammaire générique

Les relations entre *Sra*, *module\_disfluences*, et *Voice Decision* sont les mêmes que dans l'intégration précédente. La différence est, comme le montre la Figure 23, l'absence des composants de VOICE relatifs à la gestion dynamique de la grammaire d'indicatifs. De ce fait, VOICE fonctionne alors avec la grammaire générique. Pourquoi cette modification, étant donné qu'il a été prouvé que la grammaire développée est plus efficace que la grammaire générique ?

La raison est liée à l'interaction entre *module\_disfluences* et *Sra*. Dans le cas de l'utilisation de la grammaire développée, *Sra* intègre les indicatifs générés par *Autogram*. De ce fait, ce sont des indicatifs entiers qui figurent dans les énoncés envoyés à *module\_disfluences*, et non une décomposition en lettres et chiffres. Par voie de conséquence, notre agent ne peut corriger d'éventuelles disfluences intervenant lors de l'énonciation des indicatifs. En effet les disfluences interviennent justement lors de la production de l'énoncé, au niveau de ses composants élémentaires. Il est donc évident

<sup>102</sup> A chaque solution reçue par *module\_disfluences*, une nouvelle instance de celui-ci est ouverte par *IVY*. Ce détail a son importance étant donné qu'il peut affecter rapidement la rapidité d'exécution de *VOICE*, si *N* est trop élevé.

qu'aucune disfluente ne pourra être identifiée à partir d'un indicatif entier produit par *Autogram*.

Il est donc nécessaire d'utiliser la grammaire générique pour permettre à *module\_disfluences* de prendre en compte les disfluences apparaissant dans les indicatifs. Compte tenu de la corrélation du taux de disfluences avec les indicatifs, que nous avons relevé dans le chapitre 3, cette configuration de VOICE est nécessaire. Mais nous avons également choisi d'utiliser la configuration avec grammaire développée, étant donnés les résultats supérieurs de celle-ci.

#### 4.3.2. *Modification de la grammaire de reconnaissance*

Avant notre intervention, les principales caractéristiques de la CFG utilisée dans VOICE sont les suivantes :

**Tableau 29** : Caractéristiques techniques de la grammaire initiale de VOICE

Nombre de règles	Nombre d'éléments non terminaux	Nombre d'éléments terminaux
432	59	373

On observe une proportion très importante d'éléments terminaux par opposition aux éléments non terminaux. Cela est principalement dû aux règles décrivant la production des composants des indicatifs. Ces règles sont un total de 273, soit 73,19% des éléments terminaux décrits dans la grammaire initiale de VOICE.

Le tableau ci-dessous montre les caractéristiques techniques de la grammaire de VOICE après modifications de notre part. Comme on peut le voir, le nombre total de règles de réécriture a peu augmenté. Cela a été permis par la possibilité offerte par *SAPI 4* d'utiliser des symboles de type « optionnel et répétable » pour rendre compte de l'insertion éventuelle de disfluences telles que les hésitations, et de type « obligatoire et répétable » pour rendre compte des éventuelles répétitions disfluentes des éléments existants déjà dans l'ancienne version. Autrement, le nombre de règles nécessaires aurait été beaucoup plus important. La légère augmentation constatée résulte essentiellement des règles que nous avons ajoutées pour prendre en compte les hésitations et autres éditeurs, ainsi que de quelques « aménagements » nécessaires du fait de l'autre modification que nous avons opérée sur la grammaire. On observe que cette augmentation concerne les règles de production d'éléments terminaux.



**Tableau 30 :** Caractéristiques techniques de la grammaire modifiée de VOICE

Nombre de règles	Nombre d'éléments non terminaux	Nombre d'éléments terminaux
446	59	393

Cette autre modification opérée concerne l'introduction de la sémantique grâce à la grammaire. Notre objectif est d'associer à chaque mot traité une catégorie sémantico-fonctionnelle correspondante. Par exemple, indiquer que le verbe « joignez » est un ordre. Le moyen que nous avons choisi est une représentation sous la forme de paires « attribut=valeur ». Cette représentation sémantique est souvent utilisée en Traitement Automatique du Langage et en Compréhension de la Parole<sup>103</sup>.

Dans notre contexte de travail, nous choisissons de prendre comme attribut la catégorie sémantico-fonctionnelle à laquelle appartient le mot, et comme valeur le mot lui-même. Ainsi, pour reprendre l'exemple ci-dessus, le mot « joignez » sera représenté par la paire :

**Verbe\_niveau = joignez**

*SAPI 4* permet la génération de telles paires moyennant un détournement de ses mécanismes. Il est en effet possible, lors du passage de l'énoncé, de faire générer n'importe quelle chaîne de caractères lors de la reconnaissance d'un élément terminal, en plaçant entre guillemets la dite chaîne de caractères. Cette possibilité est utilisée dans VOICE pour fournir aux autres agents des informations utiles pour les traitements ultérieurs. Soit par exemple la règle :

**<Verbe\_Niveau>= <Sujet> Montez " Niveau:FL\_CLB="**

Elle impliquera la génération par la grammaire de la chaîne " Niveau:FL\_CLB=". Celle-ci sera ensuite envoyée sur IVY, pour traitement par *Voice\_Decision* qui interprétera ainsi qu'il a été fait mention d'une montée de niveau. On voit donc que des éléments sémantiques sont déjà utilisés dans VOICE au niveau de la grammaire. Par rapport à l'existant, notre approche est de systématiser l'ajout sémantique pour tout élément terminal de la grammaire (à l'exception des données numériques), et d'autre part d'utiliser une représentation « attribut=valeur »<sup>104</sup>. A titre d'exemple, voilà les règles terminales utilisées dans la nouvelle grammaire pour interpréter de l'ordre simple : « montez niveau 200 » :

**<Verbe\_expression\_Niveau> = " verbe\_Niveau=montez\_FL\_CLB " Montez**

**<mot\_niveau> = " mot\_Niveau=niveau " Niveau**

<sup>103</sup> Voir par exemple Bousquet (2002), notamment pp. 39 et 50-51.

<sup>104</sup> Du fait de ce systématisme, nous avons également modifié les représentations sémantiques incluses dans la grammaire de VOICE. De ce fait nous avons été contraints de modifier légèrement le code de *Voice\_Decision* pour l'adapter à la reconnaissance de ce nouveau formalisme.

Suite au déroulement de ces règles, la suite suivante de paires attribut=valeur sera générée :

**verbe\_Niveau=montez mot\_Niveau=niveau 200**

et envoyée sur IVY pour traitement par *module\_disfluences*.

Notons enfin qu'au niveau sémantico-fonctionnel, il est nécessaire de faire une catégorisation assez fine pour ne pas tomber sur des répétitions non disfluentes autre que pour les nombres : par exemple, il ne faudrait pas étiqueter « montez » et « niveau » par la même catégorie « *expression\_niveau* ». Autrement, l'ordre « montez niveau » serait reconnu comme une répétition alors qu'il ne l'est pas. Mais il faut également que la granularité soit suffisamment large pour ne pas avoir une série de cas particuliers. C'est ainsi qu'à toute balise correspondra une seule et unique catégorie. En tenant compte de ces principes, nous avons défini 16 attributs différents (en plus de ceux définis initialement dans la grammaire).

A la fin du processus de traitement des disfluences, *module\_disfluences* envoie à *Voice\_decision* la même chaîne de paires attributs=valeurs que celle qu'il a reçu en entrée, mais dans laquelle les paires correspondant à des éléments disfluents sont identifiées comme telles. Soit par exemple l'énoncé

**AFR1049 bonjour montez montez niveau 140**

la chaîne correspondante produite par *Sra* et reçue en entrée de *module\_disfluences* :

**Callsign=AFR1049 Politesse=CONTACT verbe\_Niveau=montez verbe\_Niveau=montez  
Niveau=FL\_CLB mot\_Niveau=niveau 140**

sera en sortie :

**Callsign=AFR1049 Politesse=CONTACT verbe\_Niveau=montez/element\_repete  
verbe\_Niveau=montez/repetition Niveau=FL\_CLB mot\_Niveau=niveau 140**

#### *4.3.3. Langage utilisé*

Comme langage de programmation de l'agent, nous avons utilisé le langage Perl. Les caractéristiques de ce langage le rendent particulièrement adapté aux nécessités de la tâche à accomplir :

- Gestion et traitement facilités et rapides des chaînes de caractères. Cela est particulièrement utile étant donné que les seuls objets manipulés par notre programme sont justement des chaînes de caractères ;
- Utilisation très puissante et souple des expressions régulières pour rechercher des patrons dans des chaînes de caractères. Par exemple, prise en compte des spécificités de la chaîne de caractères en tant que variable, définition *a priori* des « mots » (ensemble de caractères ne comprenant pas d'espace, point, etc.). Il existe

également un grand nombre de modules existants spécialisés dans le Traitement Automatique du Langage Naturel. Rappelons que pour Perl, les modules sont des ensembles de fonctions, développées par des utilisateurs du langage en dehors du cadre « officiel » de celui-ci, et offrant tout un ensemble de traitements prédéfinis. Elles jouent donc en quelque sorte le même rôle que les bibliothèques en C/C++ ;

- Manipulation dynamique des tableaux. Cette structure a été particulièrement utilisée pour le passage des énoncés ; en particulier pour l'implémentation du mécanisme de déplacement de la fenêtre ;
- Facilité d'interaction avec d'autres langages et formalismes. Cette caractéristique nous a été utile pour l'utilisation de la grammaire *CFG* dans notre algorithme ;
- Et évidemment, il fait partie des langages de programmation supportés par *IVY*. On notera d'ailleurs que la majorité des agents utilisés dans le système *VOICE* sont écrits en Perl.

Par contre, le principal inconvénient de ce langage est le fait qu'il n'est pas aussi rapide à exécuter que d'autres. Cependant, en comparaison de tous les avantages que nous venons de présenter, cela nous a semblé un faible inconvénient. Cette position est renforcée par le fait que notre but, dans le cadre de cette thèse, est de vérifier empiriquement la validité du modèle que nous proposons. Le prototype que nous proposons dans cette optique n'est donc pas contraint à répondre aux exigences habituellement requises par un système industriel ou « grand public », au premier rang desquelles la rapidité d'exécution<sup>105</sup>.

## 5. Conclusion

Dans ce chapitre, nous avons continué, développé, et achevé la description de la méthode que nous proposons pour détecter automatiquement les disfluences, commencée au chapitre 5. Nous l'avons envisagée tant du point de vue conceptuel et abstrait que de celui des détails techniques de l'implémentation.

Nous avons également présenté le système *VOICE*. Il s'agit d'un système multi-agents de compréhension automatique de la parole, spécialisé dans le dialogue de Contrôle Aérien. C'est le cadre logiciel dans lequel nous avons intégré le dispositif que nous proposons. La description que nous en avons faite porte dans un premier temps sur le fonctionnement de *VOICE* lui-même, puis sur la manière dont nous avons intégré la prise en compte des disfluences. Les deux principales modifications consistent en la création d'un nouvel agent que nous avons appelé *module\_disfluences*, et en la modification de la grammaire qui sert de moteur à la compréhension automatique.

---

<sup>105</sup> Ceci dit évidemment toutes proportions gardées ; même un prototype doit être suffisamment rapide, surtout dans le contexte du dialogue oral spontané. Mais l'exigence de rapidité est moins prégnante qu'elle pourrait l'être dans une situation réelle de contrôle aérien !

Enfin, nous avons évoqué les difficultés que nous avons rencontrées lors de l'accomplissement de cette tâche, et les conséquences qu'elles ont entraînées sur l'implémentation.

## Chapitre 7

### Validation

#### 1. Introduction

L'objectif de ce chapitre est de présenter l'ensemble du processus de validation de notre module de détection des disfluences. Nous en avons décrit dans les chapitres 5 et 7 les fondements théoriques, l'implémentation, et le taux de couverture théorique. Notre but est ici de déterminer la performance de notre module lui-même, indépendamment de celle de la reconnaissance vocale. Nous cherchons également à comprendre les causes des problèmes de performance qui pourraient se poser.

Dans un premier temps, nous présentons les fondements méthodologiques de notre processus de validation. Nous décrivons ensuite les caractéristiques du corpus de test, ainsi que le dispositif logiciel et matériel utilisé. Enfin, nous présentons les résultats obtenus, et nous les discutons, notamment à la lumière de ceux relatés par des études similaires.

#### 2. Méthodologie de validation

##### 2.1. Paramètres de la validation

Nous retenons les paramètres suivants comme pertinents pour la validation du module que nous proposons :

- Temps d'exécution : ce paramètre permet de déterminer si le dispositif que nous proposons est raisonnablement intégrable dans un système visant à fonctionner en temps réel. En effet, au-delà d'un certain seuil, la question du rapport entre le gain apporté par le dispositif et la perte qu'il occasionne en temps de calcul se poserait ;
- Taux de précision : rappelons la définition que nous donnons dans le chapitre 5, section 2.2.1 : ce taux permet d'établir le nombre de détections correctes par rapport au nombre total de détections (y compris les « faux positifs ») ;
- Taux de rappel : là encore, nous nous reportons à la définition du chapitre 5 section 2.2.1 : ce taux détermine le nombre de disfluences correctement détectées par rapport au nombre total d'occurrences ;

- Délimitation : comme nous l'indiquons au chapitre 6, section 4.2.1, la délimitation des frontières de début et de fin des disfluences est une des tâches accomplies par notre module. Il est par conséquent important de déterminer les performances dans l'accomplissement de celle-ci. Pour cette mesure également, nous utilisons les taux de précision (ici, nombre de délimitations correctes par rapport au nombre total de délimitations effectuées) et de rappel (nombre de délimitations correctes par rapport au nombre total de disfluences) ;
- Identification de la bonne catégorie : il s'agit de déterminer si une disfluence, même détectée et délimitée correctement, est également correctement catégorisée. Il peut, par exemple, arriver qu'une répétition soit étiquetée comme étant une auto-correction, ou réciproquement. Ce risque est d'autant plus important que notre algorithme de détection est basé sur l'occurrence de répétitions, que ce soit au niveau lexical ou sémantique (cf. chapitre 6, section 4.2.1, point 1). Le risque d'ambiguïté est également accru dans le cas de patrons complexes comportant des imbrications de diverses disfluences (notamment auto-corrections et répétitions). Ce paramètre concerne la justesse de l'identification des différents éléments des zones de *Repair* et de *Reparandum* : par exemple, les éléments corrigés et la correction sont-ils bien distingués les uns des autres et étiquetés comme tels ? La performance pour ce paramètre est importante : détecter et étiqueter correctement un lexème donné de la zone disfluente<sup>106</sup> est le premier pas pour pouvoir ensuite *corriger* celle-ci (par exemple, supprimer les répétitions surnuméraires, ou remplacer un mot corrigé par sa correction).

## 2.2. Constitution du corpus de validation

Dans le cadre de l'étude du dialogue oral spontané, il est difficile de se procurer des corpus. Cette difficulté est bien plus grande encore lorsque l'on s'intéresse aux disfluences dans les dialogues de contrôle aérien. Deux raisons à cela. La première est liée au contexte du contrôle aérien : compte tenu de la « sensibilité » de ce type de dialogues, pouvoir en enregistrer implique de demander des autorisations à différents interlocuteurs et organismes. Les démarches correspondantes : prises de contact, rédaction de motivations pour ces enregistrements, attente des réponses ... prennent beaucoup de temps.

D'autre part, comme nous l'avons montré dans le chapitre 3, les disfluences apparaissent beaucoup moins fréquemment dans les dialogues de contrôle aérien. Par conséquent, obtenir un nombre suffisamment représentatif de ces phénomènes dans ce contexte implique de disposer et de traiter une quantité de corpus bien plus importante que cela ne serait nécessaire pour des dialogues « quotidiens » (réservation de trains, conversations, etc.).

---

<sup>106</sup> Cf. chapitre 6, section 3.2.1, point 4.

Nous ne pouvons pas non plus utiliser les énoncés disfluents du corpus A puisqu'ils ont servi à la mise en place du module à valider. Les énoncés disfluents du corpus B, que nous n'avons pas utilisé pour la modélisation, peuvent être utilisés, et c'est d'ailleurs ce que nous faisons<sup>107</sup>. Cependant, ces énoncés sont trop peu nombreux pour une validation suffisante. C'est pour ces raisons que nous sommes contraints de produire artificiellement un corpus de disfluences, selon une méthode que nous décrivons ci-dessous.

La méthode est inspirée de la campagne « par défi » du GDR I3<sup>108</sup> (Antoine *et al.* 2002). Le but de cette campagne est de permettre de valider et d'évaluer de manière précise les systèmes de compréhension de la parole, par opposition à des bancs d'évaluation plus génériques. La méthodologie conçue dans ce cadre a été utilisée pour valider, au moins en partie, de nombreux systèmes de compréhension de la parole, par exemple Bousquet (2002), Goulian (2002), Kurdi (2003), etc.

Elle consiste à sélectionner un ensemble d'énoncés, considérés comme représentatifs du domaine de tâche étudié (Antoine *et al.* 2002, p. 563). A partir de ces énoncés de base, des énoncés *dérivés* sont construits. Le processus de dérivation consiste à transformer chaque énoncé initial de manière à ce qu'il contienne un phénomène pouvant poser problème à un système de compréhension donné<sup>109</sup>. Ainsi, pour reprendre l'exemple donné dans (Antoine *et al.* 2002, p. 563), l'énoncé initial :

non le matin à six heures environ

pourra donner l'énoncé dérivé suivant :

non c'est le matin à sept euh non à six heures environ

si le but est d'évaluer la robustesse d'un système sur le traitement des auto-corrections.

Nous avons été contraints d'opérer quelques modifications à cette méthodologie afin de l'adapter à notre contexte d'étude. En effet, le GDR I3 impliquait plusieurs chercheurs, et le corpus « dérivé » obtenu pour la campagne « par défi » est la somme des différents jeux d'énoncés conçus par chacun pour son propre système de compréhension. D'autre part, pour construire les énoncés dérivés, les participants se sont basés sur leur « intuition de locuteur natif », par exemple pour la création de disfluences. Or, cela n'est pas possible si l'on cherche à simuler les disfluences telles qu'elles apparaissent dans des dialogues de contrôle aérien. En effet, comme nous l'avons montré dans les chapitres 3 et 4, leurs modes de manifestation sont différents de ceux d'autres types de dialogues. De ce fait, il est difficile d'utiliser des règles intuitives dérivées de l'observation de dialogues non contraints. En prenant en compte ces paramètres, nous avons adapté la méthodologie « par défi » de la manière suivante :

---

<sup>107</sup> Sous certaines conditions, comme nous l'expliquons ci-dessous (section 2.3.1).

<sup>108</sup> Groupement De Recherche « Information - Interaction - Intelligence » ; <http://www.irit.fr/GDR-I3/>.

<sup>109</sup> D'où le nom d'évaluation « par défi ».

- Sélectionner un ensemble d'énoncés, ne comportant aucune disfluence, dans le corpus A<sup>110</sup>. Cet ensemble constituera la base à partir de laquelle seront dérivés les énoncés à traiter ;
- A partir de chacun de ces énoncés, construire 6 énoncés comportant chacun une disfluence. Comment générer artificiellement une disfluence compte tenu des restrictions mentionnées ci-dessus ? Nous avons choisi d'appliquer « manuellement » un des patrons que nous avons relevés, ou observés dans d'autres études. Un seul et unique patron est appliqué à un énoncé ; la position de l'énoncé à laquelle s'applique le patron est choisie aléatoirement et varie d'un énoncé à l'autre : indicatif, ordre, fréquence, etc. De même, le choix du patron à appliquer à un énoncé donné est aléatoire. Par contre, le choix des catégories de disfluences générées a été guidé par des considérations méthodologiques. Ainsi, les deux principales catégories que nous avons choisies de simuler sont les auto-corrections et les répétitions. Nous avons choisi de générer un nombre assez élevé de ces dernières par rapport à la proportion que nous avons constatée dans le corpus A<sup>111</sup>. Notre but est de compenser ainsi la sous-représentation de ce type de phénomène dans le corpus, et de déterminer si notre module est malgré tout performant dans leur traitement. Enfin, nous avons également inclut des hésitations dans certains patrons. A priori, la détection des hésitations ne devrait pas poser problème, étant donné qu'il ne s'agit pas d'un phénomène complexe à traiter. Cependant, nous avons jugé préférable de vérifier cette hypothèse, ainsi que de tester si les occurrences d'hésitations à différentes positions de l'énoncé (à l'intérieur d'indicatifs notamment) ne pose pas de problèmes de traitement. Plus généralement, nous avons divisé les énoncés dérivés en 2 groupes : le premier groupe comprend uniquement les énoncés obtenus par application des patrons que nous avons relevés sur le corpus A. Le second groupe correspond aux énoncés dérivés par application des patrons observés par d'autres auteurs : Bear *et al.* (1992), Heeman *et al.* (1994) et Kurdi (2002). Notre but est ainsi de disposer d'un maximum de patrons différents les uns des autres. D'autre part, cela permet d'éviter un éventuel biais méthodologique, qui aurait pu se présenter si l'ensemble des patrons utilisés pour générer les disfluences avaient été les mêmes que ceux à l'origine de notre module de détection des disfluences.

Voici un exemple de dérivation d'un énoncé selon la méthode décrite ci-dessus :

Enoncé initial :

F B H L B contactez ENAC 129 décimale 4

<sup>110</sup> Il aurait aussi bien pu s'agir d'énoncés issus du corpus B.

<sup>111</sup> Cf. chapitre 3, section 3.1.1.



Enoncés dérivés :

*Patrons issus du corpus A :*

F B H L B contactez **Paris 20 ENAC 129** décimale 4 (Application du patron *RIXRI*)<sup>112</sup>

F B H L B **contactez Lyon contactez ENAC 129** décimale 4 (Application du patron *MIRIMIRI*)<sup>113</sup>

I B H L B F B H L B contactez ENAC 129 décimale 4 (Application du patron *R1 M1 M2 M3 M4 / R1 M1 M2 M3 M4*)<sup>113</sup>

*Patrons observés dans d'autres études :*

F B H L B **contactez ENAC contactez le ENAC 129** décimale 4 (Application du patron *MM/MXM* (Heeman *et al.* (1994, p. 298))

F B H L B contactez ENAC 129 **décimale décimale 2 4** (Application du patron *M/MX* (Heeman *et al.* (1994, p. 298))

F B I V 2 B H L B contactez ENAC 129 décimale 4 (Application du patron *MIRIR2* (Kurdi (2002, p. 75))

### 2.3. Apparatus

Nous présentons ici les différentes données relatives, d'abord au corpus utilisé, puis au dispositif logiciel et matériel employé.

#### 2.3.1. Corpus

Le corpus que nous avons employé pour la validation est composé de plusieurs sous-ensembles correspondant chacun à un objectif d'évaluation particulier :

1. Jeux d'énoncés dérivés selon la méthode présentée en section 2.2. Nous avons sélectionné un total de 51 énoncés « de base » ; à partir de chacun de ces énoncés, 6 énoncés disfluents ont été dérivés. Un total de 306 énoncés dérivés a ainsi été mis en place.

De plus, comme nous l'avons expliqué, nous avons distingué deux jeux d'énoncés dérivés. Le critère de discrimination entre ces deux catégories est relatif à leur mode de dérivation : à partir de patrons spécifiques aux dialogues de contrôle aérien, ou d'autres études. Nous avons respecté une stricte égalité quant au nombre d'énoncés de chaque

---

<sup>112</sup> Cf. chapitre 4, section 3.3.2, tableau 27.

<sup>113</sup> Cf. chapitre 4, section 3.3.1, tableau 20.

catégorie, ce qui aboutit à un total de 153 énoncés dérivés de notre étude, et 153 énoncés dérivés d'autres études ;

2. Ensemble d'énoncés « de base » : Il s'agit des énoncés à partir desquels ont été dérivés les énoncés disfluents. La particularité de ces 51 énoncés est qu'ils ne contiennent aucune disfluence. De ce fait, il nous a paru intéressant de les faire traiter par notre module, afin de déterminer si celui-ci ne provoque pas de « faux positifs ». Il peut nous être objecté que cette mesure est incluse dans le taux de précision, et que les faux positifs peuvent également intervenir dans les énoncés comportant des disfluences. Cependant, la présence de disfluences dans l'énoncé peut provoquer des problèmes de faux positifs, par exemple dans le cas de problèmes de délimitation (les bornes de la zone de disfluence peuvent à tort inclure des composants non disfluents de l'énoncé). C'est pour éviter ce risque que nous étendons notre étude à des énoncés non disfluents ;
3. Énoncés disfluents du corpus B : Notre travail de modélisation est basé sur l'étude du seul corpus A. De ce fait, les énoncés disfluents du corpus B peuvent être inclus dans le corpus de test, étant donné qu'ils n'ont pas été utilisés pour l'élaboration du modèle. Parmi les différents sous-ensembles du corpus du test, ils présentent de plus un autre intérêt. Ce sont les seuls énoncés comportant des disfluences produites dans des situations « réelles » (par opposition à celles produites artificiellement). Dans le cadre de cette évaluation, nous avons choisi de ne traiter que les énoncés simples du corpus B, comportant des disfluences.

Les principales caractéristiques du corpus de test et des sous-ensembles qui le composent sont présentées dans le Tableau 31. On observe que le nombre de mots par énoncés (moyenne, min et max) est plus petit pour les énoncés dits « de base » que pour les énoncés dérivés. Cela n'est pas étonnant puisque la présence de répétitions ou d'auto-corrections augmente le nombre de mots de l'énoncé, parfois de beaucoup selon la longueur de la zone de disfluence. Au contraire, on remarque également que le nombre de mots des énoncés disfluents du corpus B est le plus important de tous les différents sous-ensembles. Cela est dû au fait que ces énoncés, produits dans des situations « réelles », comportent plus d'éléments que les énoncés d'apprentissage du corpus A. Et ce, même lorsqu'il ne s'agit que d'ordres simples, comme ici.

**Tableau 31 :** Caractéristiques du corpus de test

	Echantillon 1	Echantillon 2	Enoncés de base	Enoncés disfluents (corpus B)
Nombre d'énoncés	153	153	51	43
Nombre moyen de mots par énoncé	11,07	12,22	9,6	15,29
Nombre minimum de mots par énoncé	4	5	4	5
Nombre maximum de mots par énoncé	23	24	18	29

Dans ces corpus, les disfluences se répartissent de la manière représentée dans le Tableau 32, dans lequel nous donnons le nombre d'occurrences et le pourcentage correspondant (en gras). Conformément à ce que nous avons mentionné dans la section 2.2 ci-dessus, le nombre de répétitions est presque équivalent à celui des auto-corrections. Quelques occurrences d'hésitations figurent également dans le corpus.

**Tableau 32 :** Distribution des disfluences dans le corpus de test

	Echantillon 1	Echantillon 2	Enoncés disfluents (corpus B)
Auto-corrections	80 / <b>52,29%</b>	89 / <b>58,17%</b>	24 / <b>55,81%</b>
Répétitions	68 / <b>44,44%</b>	59 / <b>38,56%</b>	19 / <b>44,19%</b>
Hésitations	5 / <b>3,27%</b>	5 / <b>3,27%</b>	0 / <b>0%</b>

### 2.3.2. Dispositif employé

La validation a nécessité des modifications du module de détection des disfluences, et son intégration dans l'architecture de VOICE, décrite dans le chapitre 6. Notre validation est basée sur la représentation en chaînes de caractères des énoncés du corpus de test.

Nous avons intégré le calcul du temps d'exécution de notre module<sup>114</sup>. Afin d'obtenir une mesure du temps aussi précise que possible, nous avons utilisé le module `Perl Time::HiRes`<sup>115</sup>. Il fournit une interface à une fonction du système d'exploitation, permettant de calculer le temps écoulé à la microseconde près (autrement dit, la granularité de cette mesure est de  $10^{-6}$  secondes).

<sup>114</sup> Plus précisément, notre but étant de mesurer le temps d'exécution du seul algorithme de traitement des disfluences, nous avons calculé le temps au début et à la fin du code correspondant. Nous avons ainsi ignoré les portions de code relatives à d'autres aspects du programme : ouverture, écriture, et fermeture de fichiers, affichage à l'écran, etc.

<sup>115</sup> <http://search.cpan.org/dist/Time-HiRes/HiRes.pm>

Enfin, nous n'avons pu valider la partie de notre algorithme relative à l'analyse générique robuste de patrons complexes (décrite au chapitre 6, section 4.2.1, point 4). Pour compenser son absence, nous avons augmenté le nombre de patrons recherchés directement par l'algorithme.

Sur le plan matériel, la machine sur laquelle le corpus a été traité est équipée d'un processeur monocœur AMD© 64 3000+ et d'1,2 Go de mémoire vive.

### 3. Résultats de la validation

Nous présentons dans cette section les résultats que nous avons obtenus. Les paramètres que nous avons décrits en section 2.1 sont regroupés dans deux sous-parties : le temps d'exécution d'une part, et les performances de détection et de traitement (catégorisation, délimitation) d'autre part. Enfin, nous proposons une analyse complémentaire de ces résultats dans un troisième temps.

#### 3.1. Temps d'exécution

Nous avons calculé le temps moyen d'exécution pour les échantillons 1 et 2. Les résultats sont donnés dans les Tableau 33 et Tableau 34 ci-dessous. On observe que le temps d'exécution est rapide puisqu'il est en moyenne inférieur à 2 millisecondes (ms).

**Tableau 33 :** Temps d'exécution (échantillon 1)

Temps Moyen	1,48
Temps Min	0,16
Temps Max	3,61

**Tableau 34 :** Temps d'exécution (échantillon 2)

Temps Moyen	1,76
Temps Min	0,18
Temps Max	5,18

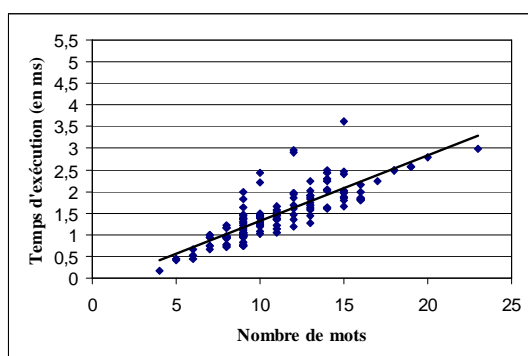
On constate également que le temps d'exécution est plus long pour l'échantillon 2. Nous voyons deux explications à cela. D'une part le fait que les énoncés comportent plus de mots (cf. ci-dessous). D'autre part, ils ont été formés à partir de patrons plus complexes : nombre de « répétables » et/ou de « répétés »<sup>116</sup> supérieur à 4, intrication de plusieurs corrections et répétitions, etc. Ils nécessitent ainsi des traitements plus complexes et donc plus longs : recherche exhaustive dans la liste des patrons modélisés, application de plusieurs règles de traitement, etc.

A titre de comparaison, Truillet *et al.* (2004, p. 15-16) rapportent un temps moyen de 21 à 30 ms pour la reconnaissance et la compréhension automatique d'indicatifs prononcés oralement. Ce temps rapporté est plus important étant donné qu'il intègre aussi le temps de la reconnaissance du signal vocal ; de plus cette évaluation a été faite sur une machine moins puissante.

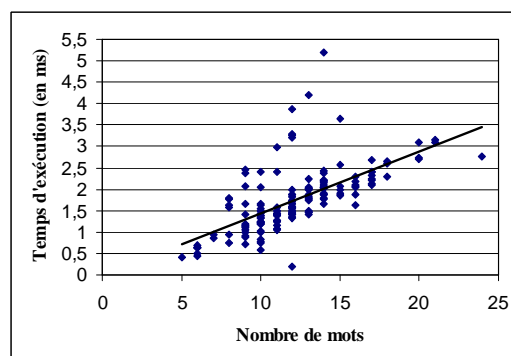
---

<sup>116</sup> Cf. Chapitre 1, section 3.2.4.4.

D'autre part, la corrélation entre le temps d'exécution et le nombre de mots est avérée : les nuages de points représentés dans les figures ci-dessous indiquent clairement que le temps de traitement est fonction du nombre de mots traités. On remarque quelques déviations par rapport à la courbe de tendance. Nous les attribuons au fait que les énoncés « déviants » comportent des disfluences complexes. De ce fait, ils ont nécessité un traitement plus complexe (au sens vu ci-dessus) par le module, qui se traduit lui-même par un temps supérieur. Cette interprétation est renforcée par le fait que les déviations sont plus nombreuses pour l'échantillon 2 : or, on sait que ce dernier comporte des patrons plus complexes que ceux de l'échantillon 1, puisqu'ils sont issus de l'observation de corpus non contraints<sup>117</sup>.



**Figure 24 :** Rapport entre le nombre de mots et le temps d'exécution (échantillon 1)



**Figure 25 :** Rapport entre le nombre de mots et le temps d'exécution (échantillon 2)

En conclusion, ces résultats montrent que le traitement automatique des disfluences peut être intégré à une chaîne complète de traitement automatique de la parole, sans alourdir le temps global de traitement.

### 3.2. Performances du module

Les résultats varient selon le sous-ensemble de corpus traité, le type de disfluences rencontrées, etc. En conséquence, nous organisons cette présentation en trois sous-parties. La première est consacrée à un aperçu général des performances observées ; nous y montrons également les résultats obtenus avec les occurrences d'hésitation, ainsi que le taux de catégorisation correcte des disfluences. Les deux sous-parties qui suivent sont consacrées à l'examen détaillé des résultats obtenus respectivement pour la détection et la délimitation des auto-corrections et des répétitions.

#### 3.2.1. Aperçu général

Le Tableau 35 ci-dessous présente les résultats obtenus pour trois des sous-ensembles du corpus de test.

<sup>117</sup> Cf. les comparaisons que nous avons effectuées au chapitre 4, section 4.

**Tableau 35 : Performances globales**

	Échantillon 1		Échantillon 2		Échantillon corpus B	
	Détection générale	Délimitation	Détection générale	Délimitation	Détection générale	Délimitation
Nombre de détections correctes	142	129	139	87	36	35
Nombre total de détections	161	148	155	103	38	43
Nombre total de disfluences	153	153	153	153	43	43
Taux de précision	88,20%	87,16%	89,68%	84,47%	94,74%	89,74%
Taux de rappel	92,81%	84,31%	90,85%	56,86%	83,72%	81,40%

Ce tableau ne comprend pas les résultats obtenus avec le corpus d'énoncés non-disfluents. En effet, une seule mesure concernant ce corpus est permanente : le très faible taux de faux-positifs (moins de 6%). Tous sont causés par des problèmes d'identification des balises ou de valeurs numériques (cf. section 3.3.1 ci-dessous). Ce résultat, associé aux taux de précision élevés du Tableau 35, indique que notre module prend bien en compte les spécificités du contrôle aérien (indicatifs, balises) et ne produit que peu de faux-positifs.

Les taux de performance pour les hésitations sont par contre inclus dans ce tableau. Les taux de précision et de rappel pour cette disfluence sont de 100%, et aucun faux-positif ou autre problème n'est à signaler.

Plus généralement, on observe que les meilleures performances sont obtenues avec l'analyse de l'échantillon 1.

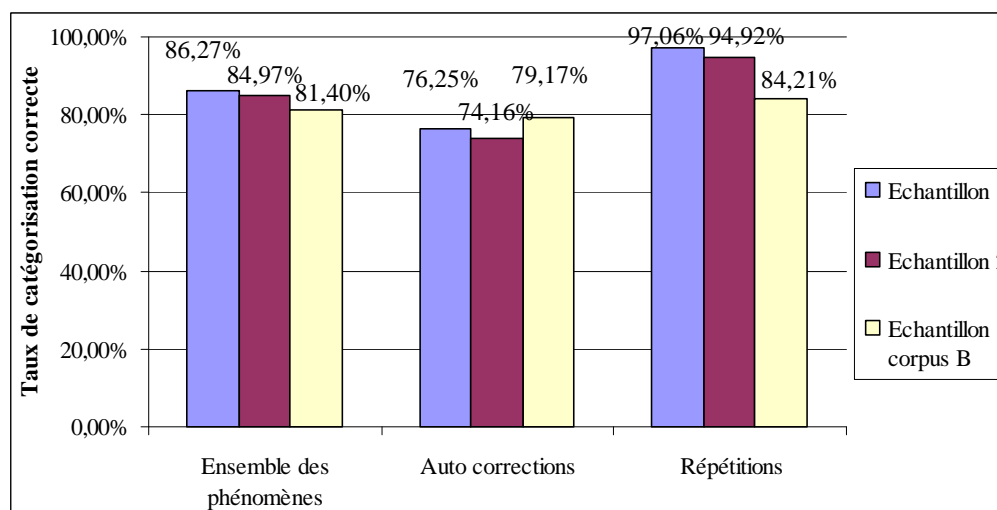
Au contraire, l'échantillon 2 présente des résultats plus faibles. Ils sont malgré tout corrects, compte tenu du fait qu'ils résultent de patrons que nous n'avons pas modélisés. Cependant, le pourcentage particulièrement faible du taux de rappel pour les délimitations (moins de 57%) montre que les patrons non modélisés ont malgré tout posé des problèmes à notre module. Globalement, celui-ci détecte bien la présence d'une disfluence, la catégorise correctement, mais rencontre des difficultés pour en cerner les limites, surtout au niveau de la frontière de fin de la zone. Nous expliquons ce problème par le fait que des disfluences de ce sous-ensemble ont été générées à partir de patrons très complexes :

nombre de « répétables » et/ou de « répétés »<sup>118</sup> supérieur à 4, intrication de plusieurs corrections et répétitions, etc.

Notre module obtient également de bonnes performances sur l'échantillon issu du corpus B. Ces derniers résultats sont cependant à considérer avec précaution, étant donnée la petite taille de ce sous-ensemble.

Dans les deux sous-sections qui suivent, nous décrivons plus en détail ces résultats, en les analysant selon qu'ils concernent les auto-corrections ou les répétitions.

Auparavant, examinons la performance dans la détection de la catégorie de disfluence (en section 2.1), et l'étiquetage correspondant des lexèmes de la zone disfluente. Les résultats sont représentés par la Figure 26, avec précision des corpus utilisés et des phénomènes impliqués (ensemble, auto-corrections, répétitions). Globalement, les résultats sont bons. Les meilleurs scores sont obtenus au niveau du traitement des répétitions. Ce n'est guère étonnant puisqu'il s'agit d'un phénomène plus simple à traiter que les auto-corrections. Le pourcentage obtenu pour les répétitions avec l'échantillon issu du corpus B peut surprendre, car il est plus bas de 10% que les deux autres. Après analyse, il apparaît que ce résultat est dû à un problème de représentation dans la grammaire des éléments répétés, concernant deux énoncés (sur 19). Le mécanisme de catégorisation lui-même n'est par conséquent pas en cause.



**Figure 26 :** Taux de catégorisation correcte des disfluences par notre module

#### 3.2.2. Auto-corrections

Globalement, les performances diffèrent peu selon les corpus. L'échantillon du corpus B obtient un score de rappel particulièrement élevé ; cependant ce résultat est à tempérer du fait de la faible représentativité des énoncés considérés (24 au total).

<sup>118</sup> Cf. Chapitre 4, section 3.2.4.4.

**Tableau 36 :** Performances de traitement des auto-corrections

	Échantillon 1		Échantillon 2		Échantillon corpus B	
	Détection générale	Délimitation	Détection générale	Délimitation	Détection générale	Délimitation
Nombre de détections correctes	70	61	78	43	19	19
Nombre total de détections	78	69	88	46	20	21
Nombre total de disfluences	80	80	89	89	24	24
Taux de précision	89,74%	88,41%	88,64%	93,48%	95,00%	90,48%
Taux de rappel	87,50%	76,25%	87,64%	48,31%	79,17%	79,17%

### 3.2.3. Répétitions

**Tableau 37 :** Performances de traitement des répétitions

	Échantillon 1		Échantillon 2		Échantillon corpus B	
	Détection générale	Délimitation	Détection générale	Délimitation	Détection générale	Délimitation
Nombre de détections correctes	67	63	56	39	19	16
Nombre total de détections	78	78	62	42	20	18
Nombre total de disfluences	68	68	59	59	24	19
Taux de précision	85,90%	80,77%	90,32%	92,86%	95,00%	88,89%
Taux de rappel	98,53%	92,65%	94,92%	66,10%	79,17%	84,21%

Les performances sont globalement bien meilleures pour les répétitions que pour les auto-corrections. L'explication donnée en section 3.2.1 s'applique également ici : les répétitions, surtout dans le cadre contraint du trafic aérien sont moins complexes à traiter que les auto-corrections, que ce soit en termes de structuration de la zone disfluente ou de variété de patrons.



#### 3.3. Analyses complémentaires

Dans la section précédente, nous avons analysé et commenté les résultats obtenus, de manière à les expliquer. Nous allons maintenant apporter deux nouveaux éclairages sur ces résultats. Dans un premier temps, nous cherchons à déterminer les causes des erreurs. Dans un deuxième temps, nous situons notre validation par rapport à d'autres travaux similaires.

##### 3.3.1. Cause des erreurs

Nous avons examiné chacune des erreurs de traitement (détection, délimitation, catégorisation) afin de comprendre leur origine, et le cas échéant proposer des solutions pour l'améliorer. Nous avons ainsi identifié 6 principales causes à l'origine de ces erreurs de traitement. Ces causes sont :

- Identification des valeurs numériques : désigne toute erreur générée par un problème d'identification des valeurs numériques telles que les indicatifs, niveaux, ou fréquences. Rappelons<sup>119</sup> que nous avons mis en place, dans notre module, des règles dont le rôle est de gérer ces valeurs, afin d'éviter les faux positifs. Plus précisément, ces erreurs sont principalement liées aux faux positifs et aux problèmes de délimitation ;
- Gestion des balises : le mot « Balise » est ici à prendre au sens utilisé dans la phraséologie<sup>119</sup> : il s'agit d'une étape inscrite dans un plan de vol. Comme pour les valeurs numériques, nous avons mis en place des règles de gestion des balises<sup>119</sup>. Les erreurs provoquées par une insuffisance au niveau de ce dispositif appartiennent à la catégorie « gestion des balises » ;
- Complexité de la disfluence : ces erreurs sont provoquées par une complexité importante de la zone disfluente, par exemple l'intrication de plusieurs corrections et répétitions ;
- Problème de grammaire : les problèmes appartenant à cette catégorie sont liés à la manière dont les composants d'un énoncé sont représentés par la grammaire CFG décrites au chapitre 6, sections 3.2.3 et 4.3.2. Soit, par exemple, l'énoncé :

contactez      De      De      Gaulle      121,15      au      revoir

La disfluence y est une répétition qui affecte un composant de « De Gaulle ». Or, la représentation dans la grammaire de cet élément de l'énoncé est « Contact=Degaulle ». Une telle représentation ne permet pas de rendre compte de la disfluence. Deux solutions sont envisageables : d'une part, introduire des modifications dans la grammaire, lorsque cela est possible. D'autre part, lorsque de telles

---

<sup>119</sup> Cf. chapitre 6, section 3.2.1, point 2.

modifications sont difficiles (comme dans notre exemple), introduire des règles, en aval de la grammaire, pour prendre en compte ces cas particuliers. Dans cette même catégorie, nous avons également relevé des problèmes liés à la traduction d'un terme anglais en français, ou inversement ;

- Principe de base de notre algorithme de décodage : les erreurs de ce type sont les plus problématiques. En effet, il est difficile d'améliorer ce type de cause à moins de changer en profondeur l'algorithme. Elles ne concernent que les auto-corrections. Appartiennent à cette catégorie deux types de causes. La plus fréquente est liée au dispositif de détection d'une auto-correction, basé sur l'occurrence d'une répétition au niveau sémantico-fonctionnel. Il arrive cependant que des auto-corrections ne se manifestent par aucune répétition à ce niveau. Par exemple, l'énoncé

Air Vendee 513 H F descendez niveau contactez ENAC 129 4 au revoir

comporte une auto-correction dans laquelle « descendez niveau » est remplacé par « contactez ENAC ». Notre module ne peut la détecter, car les deux verbes n'appartiennent pas à la même catégorie sémantique (représentée dans la grammaire par l'attribut de la paire attribut-valeur). L'autre type de cause est une distance entre les répétitions supérieure à 10 éléments : dans ce cas, le phénomène n'a pas été détecté dans le cadre de la fenêtre, limitée à 10 éléments maximum. Ce cas ne concerne cependant qu'un seul énoncé de l'ensemble du corpus de test ;

- Inconnu : cette catégorie regroupe les erreurs dont nous n'avons pas pu identifier les causes.

Le Tableau 38 indique la distribution, en pourcentage et par ordre décroissant, de ces causes, dans l'ensemble du corpus. Les différences entre les sous-ensembles du corpus ne sont pas assez pertinentes pour nécessiter une présentation distincte par sous-ensemble.

**Tableau 38** : Pourcentage des différentes causes d'erreurs de traitement

Cause de l'erreur	Pourcentage
Problème de grammaire	37,89%
Complexité de la disfluence	26,09%
Identification des balises	14,29%
Identification des valeurs numériques	9,32%
Principe de l'algorithme	8,70%
Inconnu	3,71%

Cette distribution constitue elle aussi une validation de notre module. En effet, la quasi-totalité des erreurs relèvent d'une cause qui peut être corrigée par un affinement des règles de traitement ou de la grammaire. Il est donc fortement probable que nous puissions augmenter encore les performances de notre module. Deux exceptions : les problèmes liés directement au mode de détection de notre algorithme, et ceux dont nous n'avons pas identifié la cause. En ce qui concerne ces derniers, il est vraisemblable qu'une analyse plus détaillée permettra de les identifier.

Par contre, il est plus difficile de remédier aux problèmes liés directement à notre algorithme. Une solution envisageable serait d'identifier d'autres indices que la détection de répétitions au niveau lexical ou sémantique ; par exemple la présence de marqueurs lexico-syntaxiques (tels que ceux présentés dans le chapitre 3, section 2.3 et *sq.*) l'incomplétude d'un ordre, etc. Cette solution entraîne cependant deux inconvénients : d'une part, un accroissement du temps d'exécution ; d'autre part, une augmentation probable des faux positifs et donc une baisse correspondante du taux de précision.

#### 3.3.2. Comparaison avec d'autres études

Il est difficile de mener une comparaison objective avec les résultats d'autres travaux sur le traitement automatique des disfluences, et ce pour diverses raisons. La principale est que nous avons effectué une validation, et non une évaluation. De plus, comme nous le mentionnons dans le chapitre 1 (section 3.2.1) la terminologie varie d'un système à l'autre. Cela peut compromettre l'exactitude d'une comparaison inter-étude. Il est malgré tout intéressant de nous situer par rapport aux études dont les plus proches de la nôtre.

Nous regroupons dans un tableau synoptique les résultats obtenus par les études similaires. Nous ne prenons en compte ici que les performances mesurées en termes de taux de précision et de rappel, afin de conserver une métrique commune.

**Tableau 39 :** Présentation synoptique des performances d'autres systèmes de détection/traitement des disfluences

Auteurs	Heeman <i>et al.</i> (1994, p. 300)		Kurdi (2002, p. 171)				Johnson <i>et al.</i> (2004b, p. 38)	Liu (2003, p. 40)
Type de disfluence	Auto-correction		Répétitions		Auto-correction		Mots situés dans le <i>Reparandum</i>	Fragments de mots
	Détection	Correction	Détection	Délimitation	Détection	Délimitation		
Taux de précision	89%	86%	98,75%	90%	77,55%	85,63%	79% <sup>120</sup>	74,3%
Taux de rappel	83%	80%	96,20%	91,13%	92,68%	71,42%	75,90% <sup>120</sup>	70,1%

<sup>120</sup> Moyenne de plusieurs résultats

On observe que les résultats sont assez hétérogènes en fonction des études, des types de disfluences, et de la mesure (précision ou rappel). Cependant, aucun taux n'est inférieur à 70%. En comparant avec nos résultats (avec les précautions méthodologiques évoquées ci-dessus), il apparaît que notre module se situe dans la moyenne supérieure des performances observées. La seule exception concerne le taux de rappel obtenu, pour la délimitation avec l'échantillon 2. Mais nous avons montré que cela est principalement dû à l'utilisation de patrons de disfluences non modélisés, car apparaissant dans des productions orales non contraintes.

#### 4. Conclusion

Nous avons présenté dans ce chapitre le processus de validation de notre système. Il consiste, dans un premier temps, à définir une méthodologie.

L'objectif de celle-ci est de déterminer les paramètres à valider, et le corpus de test à utiliser. Les paramètres retenus sont le temps d'exécution, les taux de précision et de rappel pour la détection et la délimitation, et le taux de catégorisation des disfluences. Le corpus de test est obtenu en générant des énoncés non disfluents à partir d'un ensemble d'énoncés de base, selon une méthodologie décrite dans Antoine *et al.* (2002), et adaptée au contexte de notre étude.

Les résultats obtenus sont très satisfaisants. Ils indiquent en effet que le temps d'exécution du module est très rapide, et laisse donc de la marge pour une intégration à un système plus complexe (tel que VOICE), ou l'ajout de règles supplémentaires. On observe également que le module est robuste, et bien adapté à la tâche et le contexte pour lequel il a été conçu.

Enfin, cette validation nous a également permis de mettre en évidence les lacunes de notre module, et d'envisager les moyens d'y remédier. Il reste cependant une évaluation à mener. Nous revenons sur ces points en conclusion et dans les perspectives.

## Conclusion

Ce travail a été consacré à l'étude des disfluences dans un cadre particulier : le contrôle du trafic aérien. C'est à notre connaissance le premier travail de ce type, du moins dans la sphère francophone. Dans cette conclusion, nous présentons les contributions apportées par notre travail, ainsi que les perspectives que nous envisageons.

### Contributions

Dans un premier temps, nous avons effectué un état de l'art détaillé consacré à notre contexte de travail : les disfluences d'une part, et le travail de contrôle aérien d'autre part. L'analyse de la littérature sur les disfluences nous montre que la catégorisation et la dénomination de ces phénomènes est difficile et varie d'un auteur à l'autre. Nous avons, malgré tout, dégagé un ensemble de caractéristiques communes, aux causes de l'apparition des disfluences, de leurs différents modes de manifestations, et des conséquences sur le reste de la production orale dans laquelle elles s'insèrent. Nous nous sommes intéressés également à l'aspect cognitif sous-jacent de ces phénomènes.

Nous avons ensuite présenté le contexte du Contrôle du Trafic Aérien. D'un point de vue linguistique, la principale caractéristique à retenir est l'utilisation de la phraséologie. Celle-ci concerne aussi bien le lexique utilisé que la syntaxe employée. Elle a pour unique fonction de permettre la réalisation de la tâche de contrôle, en prenant en compte les impératifs de la tâche, parmi lesquels les plus importants sont la rapidité d'exécution et l'intelligibilité. De ce fait, les schémas lexico-syntaxiques présentent assez peu de variété, et sont particulièrement contraints par la tâche.

Nous avons démontré que la particularité de la tâche de contrôle aérien a également de l'influence au niveau cognitif, puisqu'elle induit une très forte charge cognitive. Celle-ci est par contre compensée par le relativement faible nombre de schémas à maîtriser. Cette dernière remarque est surtout valable pour l'utilisation de la phraséologie en situation « réelle » mais sans doute moins pour celle de l'apprentissage des contrôleurs aériens. L'étude de celle-ci présentant de nombreux intérêts, nous en avons présenté les principales caractéristiques, ainsi que la manière dont l'apprentissage est réalisé.

Ma contribution a d'abord consisté en une analyse linguistique fine et détaillée des deux corpus, de grande taille :

- Corpus A : dialogues d'apprentissage entre des contrôleurs aériens en formation et des « pseudos-pilotes ». Par rapport à la phraséologie et ses caractéristiques, ces dialogues sont identiques à ceux récoltés dans des conditions de trafic réel. En effet, comme nous l'avons décrit dans le chapitre 2, section 3.3, ces sessions d'apprentissage sont aussi proches que possible des conditions réelles. Les dialogues qui en résultent sont donc représentatifs de l'utilisation de la phraséologie. Notre connaissance de ce corpus, et son analyse ultérieure, a été accrue par le fait que nous en avons nous-mêmes effectué la transcription intégrale ;
- Corpus B : dialogues de contrôle aérien, issus de situations réelles (centre de contrôle « en route », cf. 4.1.1.2). Les transcriptions ont été assurées par un organisme tiers.

J'ai conduit une analyse exhaustive de tous les aspects pertinents des disfluences rencontrées : catégorisation, modes de manifestation, distribution, contexte d'apparition. L'objectif étant de vérifier l'hypothèse formulée en introduction, selon laquelle les disfluences apparaissant dans le cadre de dialogues de Contrôle Aérien présenteraient des spécificités. Le résultat est la disponibilité d'une expertise des types de disfluences dans le contrôle aérien.

L'hypothèse s'avère confirmée. L'un des résultats les plus remarquables que nous avons mis en valeur est la grande différence qui existe entre la fréquence des disfluences et leur mode d'apparition constatées dans des corpus de contrôle de trafic aérien en comparaison avec celles de tout autre corpus : il y a beaucoup moins de disfluences lorsque la phraséologie est utilisée.

Cette différence est vraie quel que soit le type de disfluence considéré. Nous constatons par contre que les disfluences apparaissent plus fréquemment dans le contexte de certains éléments de la phraséologie. Nous avons classifié ces éléments, non pas en fonction de critères lexico-syntaxiques comme cela se fait usuellement dans la littérature, mais selon des critères fonctionnels. Cette démarche nous est apparue plus appropriée aux caractéristiques linguistiques induites par la phraséologie. Nous avons aussi établi un lien entre l'utilisation des marqueurs lexicaux et la phraséologie qui dirige normalement la production des énoncés dans le contexte du Contrôle de Trafic Aérien.

Toujours dans le cadre d'une analyse exhaustive, nous avons pris en compte deux niveaux qui sous-tendent toute communication humaine : linguistique (chapitre 3, section 5.1) et cognitifs (section 5.2). Il s'agit d'une démarche originale dans un travail à dominante informatique. Ces deux points de vue nous ont permis de proposer des explications à la spécificité du corpus quant aux disfluences, en dehors de la seule influence de la phraséologie.

Ma deuxième contribution consiste en la conception d'un dispositif de détection automatique des disfluences dans le cadre particulier de la phraséologie. Cette phase de conception comprend plusieurs étapes, allant de la modélisation théorique à l'implémentation du module logiciel, pour finir par l'évaluation de ce dernier.

Nous avons d'abord montré que les seuls marqueurs lexicaux-syntaxiques ne sont pas suffisants pour détecter les disfluences, et qu'il est nécessaire de passer à un niveau supérieur d'abstraction. Pour cela, nous avons adopté un schéma de modélisation mis au point par Bear *et al.* (1993). Ses principaux intérêts sont l'adaptation à l'objet d'étude (les disfluences) et une grande capacité d'expression et de représentation. J'ai contribué à l'extension de ce schéma en rajoutant des marques d'annotation adaptées à notre étude.

Nous avons mis en place une méthodologie de traitement automatique des disfluences à partir des observations de la modélisation, et d'un état de l'art sur les techniques existantes. Cette méthodologie repose sur la détection de patrons, correspondant à différentes catégories de disfluences, et est complétée par la représentation de chaque élément de l'énoncé traité sous la forme d'une paire « attribut=valeur ». A partir de l'état de l'art des méthodes existantes, et de l'analyse des modèles de disfluences de notre corpus, nous avons démontré que cette approche est la plus adaptée aux spécificités entraînées par la phraséologie du contrôle aérien. L'algorithme décrivant cette procédure, et l'implémentation qui en résulte, ont été parfaitement intégrés dans le système de compréhension automatique d'ordres simples de contrôle aérien, appelé VOICE.

Enfin, nous avons validé le module d'identification des disfluences que nous avons conçu. Dans un premier temps, nous avons mis en place une méthodologie de validation, qui réponde aux objectifs suivants : mesurer les performances, et comprendre les causes d'erreurs. Pour cela, nous avons construit un corpus de tests de 400 énoncés, composé de quatre différents sous-ensembles. Il apparaît que notre module obtient de bonnes, voire très bonnes performances, et est robuste. Notre hypothèse sur la spécificité du traitement automatique des disfluences dans la phraséologie est confirmée elle aussi. Il montre également que la technique des patrons est très bien adaptée, aussi bien à la modélisation qu'au traitement automatique des disfluences dans un langage contraint.

### **Perspectives**

A très court terme, notre premier objectif est de fournir une évaluation du module que nous avons implémenté. Cette évaluation aura deux objectifs. D'une part, mesurer l'impact des améliorations du module, effectuées à partir d'observations relevées lors de la validation. D'autre part, prendre en compte l'intégration du module dans le système VOICE, et déterminer l'influence de l'aspect vocal.



## Conclusion

---

L'évaluation se fera avec la collaboration du DSNA et de l'ENAC. Elle portera sur deux corpus différents :

- Le corpus A<sup>121</sup> : Utiliser ce corpus pour une évaluation peut sembler étonnant étant donné qu'il nous a servi à concevoir notre modèle. La principale utilité est de déterminer l'influence du moteur de traitement du signal sur les résultats obtenus. En effet, comme l'ont montré les études des projets *PAROLE*, *VOCALISES*, et *VOICE* (citées dans le chapitre 3, section 4.1.1.1), les performances de tout système de compréhension de la parole sont dépendantes de celles du moteur de reconnaissance du signal. Par contre, nous n'utiliserons dans cette session d'évaluation que les seuls énoncés comportant au moins une disfluence ;
- Un nouveau corpus (appelé corpus C) : Ce corpus va être enregistré lors de sessions de pseudo-pilotage, dans les mêmes conditions que celles décrites dans le chapitre 2. Un total d'environ 2 heures d'enregistrements (sans pause) est prévu. Nous transcrivons avec Transcriber, et selon les conventions décrites dans le chapitre 2, les énoncés de ce corpus comportant au moins une disfluence. Les possibilités d'alignement sur le signal offertes par Transcriber permettront ainsi de savoir à quelle position exacte de l'enregistrement se situe chaque disfluence.

Comme nous l'avons signalé dans le chapitre 6, une des limites de *VOICE* est l'impossibilité de traiter des ordres complexes. Pour « contourner » ce problème sans avoir à recueillir un corpus composé uniquement d'ordres simples, un pré-traitement des corpus s'impose. Il consiste à segmenter manuellement, au moyen d'un logiciel d'édition audio<sup>122</sup>, les énoncés complexes, afin d'obtenir une série d'ordres simples. Les fichiers correspondants seront ensuite fournis en entrée à *VOICE*. Cette procédure n'est pas sans poser quelques problèmes d'ordre méthodologiques : par exemple la segmentation peut dénaturer le sens de certains énoncés complexes. Cependant, au vu de notre analyse des corpus A et B, nous estimons que ce risque potentiel ne concernerait que peu d'énoncés.

Le but de cette évaluation sera de comparer la transcription de chaque énoncé disfluent avec la production générée par notre dispositif. L'objectif de cette comparaison sera de déterminer les paramètres suivants :

- Taux de rappel et taux de précision (tels que définis dans le chapitre 5, section 2.2.1) : ces mesures sont un indicateur reconnu des performances du dispositif considéré ;

---

<sup>121</sup> Le corpus B est inexploitable par un moteur de traitement du signal, compte tenu de sa qualité audio (cf. chapitre 3).

<sup>122</sup> *Audacity* : <http://audacity.sourceforge.net/> .

- Temps mis pour traiter les énoncés : cette mesure n'est pas la plus importante, puisqu'on ne considère ici qu'un prototype. Elle est cependant intéressante pour des utilisations (et améliorations) ultérieures en environnement « réel » ;
- Cause des erreurs : pour chaque fausse interprétation de notre module, nous déterminerons quelle est la cause de l'erreur : problème de reconnaissance de l'énoncé, problème lié à notre algorithme, autre (cf. chapitre 7, section 3.3.1) ;
- Influence des  $n$  meilleures solutions : les trois paramètres décrits ci-dessus seront évalués selon trois cas de figure, différents selon le nombre de solutions données par le module *Sra* (cf. chapitre 6, section 3.2.3) : respectivement la meilleure solution, les trois meilleures, et les dix meilleures. Pourquoi ce paramètre supplémentaire ? On sait que la première solution donnée par le module de reconnaissance n'est pas systématiquement celle qui correspond le mieux à ce qui a été prononcé. C'est pourquoi il est nécessaire de déterminer si disposer de plus de solutions améliore les performances de notre module de détection et traitement des disfluences. La prise en compte de ce paramètre est classiquement utilisée dans les évaluations de systèmes de Compréhension Automatique de la Parole (cf. par exemple Bousquet (2002, chapitre 4)).

En donnant en entrée à VOICE l'ensemble des énoncés, même non disfluents, du corpus C, nous pourrions également déterminer le taux de surgénération (production de « faux positifs ») de notre algorithme. Dans le cas de « faux positifs » sur des énoncés non disfluents, nous transcrivons l'énoncé correspondant afin de pouvoir analyser finement la ou les causes d'erreurs.

A moyen terme, notre but est d'améliorer le système que nous avons décrit dans les chapitres 5 et 6. Les améliorations concernent respectivement le système VOICE dans son ensemble, et notre module en particulier.

La première est de recourir à un moteur de reconnaissance plus récent et performant. Le système VOICE utilise un moteur de reconnaissance vocale datant de plusieurs années. Compte tenu des progrès accomplis dans ce domaine, il paraît évident, avant même évaluation, qu'utiliser une version plus récente augmenterait les performances du système.

Une autre amélioration possible de VOICE consisterait à prendre en compte les ordres complexes, en plus des ordres simples. Les possibilités d'expression du système seraient ainsi étendues à tout type de production orale de Contrôle Aérien. Cela nécessiterait néanmoins une réécriture quasi-totale de la grammaire utilisée.

En ce qui concerne notre module, la principale amélioration consisterait à augmenter le nombre de disfluences détectées. Nous pensons, notamment, à l'impossibilité actuelle de détecter les amorces. Cette impossibilité est induite par des limitations des systèmes de

## Conclusion

---

traitement du signal, comme nous l'avons expliqué dans le chapitre 6. Si cette limitation disparaissait, nous pourrions prendre en compte également les amorces, et les amorces correctives.

En effet, nous avons vu dans le chapitre 4 que les amorces fonctionnaient sur le même schéma que les auto-corrections, à l'exception de la prononciation de l'élément corrigé, qui est interrompue avant la fin. Or, du fait de cette exception, il est impossible de disposer de l'un des éléments de la fenêtre : celui qui détermine la borne gauche de celle-ci. Or, nous avons constaté que les amorces sont souvent reconnaissables dans le contexte du contrôle aérien (cf. chapitre 4). Cela est dû à la phraséologie, qui restreint les ambiguïtés entre plusieurs choix possibles.

Nous proposons par conséquent la solution suivante pour traiter les amorces (correctives ou non). Elle implique au préalable de disposer d'une base de données comportant l'ensemble du lexique sous forme décomposée en phonèmes (comme cela est fait avec la base *BDLex*, cf. De Calmès *et al.* 1998). Cette base constituée, il s'agit ensuite d'obtenir la représentation de mot « inconnu »<sup>123</sup>. Pour cela l'utilisation d'un phonétiseur<sup>124</sup> est nécessaire. Enfin, il faut comparer la représentation obtenue avec les éléments de la base de données. Pour cette tâche, l'utilisation de la mesure de Levenshtein (Levenshtein, 1966) s'impose : elle permet d'attribuer un score en fonction de la différence entre deux chaînes de caractères (ou de phonèmes, en l'occurrence)<sup>125</sup>.

De ce fait, il faut introduire la modification suivante dans l'algorithme :

```
Si un mot inconnu est parsé
  Interrompre l'algorithme
  Phonétiser le mot
  Comparer le résultat de la phonétisation avec la
  base de données
  Si au moins un mot correspond
    Sélectionner le candidat avec le score le
    plus élevé
    L'insérer dans l'énoncé
    Le marquer comme « amorces »
  Sinon
    Marquer le mot comme « fragment de
    mot »
Finsi
Finsi
```

---

<sup>123</sup> Dans le contexte de l'implémentation, un mot inconnu est un mot n'appartenant pas à la grammaire.

<sup>124</sup> Par exemple *LIA\_PHON* (développé par Frédéric Béchet : [http://lia.univ-avignon.fr/fileadmin/documents/Users/Intranet/chercheurs/bechet/download\\_fred.html](http://lia.univ-avignon.fr/fileadmin/documents/Users/Intranet/chercheurs/bechet/download_fred.html)), qui convertit un texte en la suite de phonèmes correspondants.

<sup>125</sup> Pour un exemple d'utilisation de cette mesure, voir par exemple Truillet *et al.* (2005 p. 69-70).

A plus long terme, nous envisageons trois axes de perspectives pour notre travail, selon les disciplines concernées :

- Interaction Homme-Machine (IHM) : en termes d'IHM, la principale problématique est le mode d'intégration de notre dispositif à l'interface de pseudo-pilotage que nous avons décrit dans le chapitre 2. L'intérêt de cette intégration est de signaler aux opérateurs humains l'occurrence d'un problème qu'il pourrait ne pas remarquer. Ce type d'indication peut améliorer la robustesse de l'interaction homme-machine médiée (cf. Deherty *et al.* 2005, particulièrement p. 6-7) ;
- Sciences cognitives : deux principaux axes d'étude sont relatifs à ce domaine.
  - Le premier axe est relié à l'intégration, que nous proposons ci-dessus, dans une IHM. Il s'agirait de mesurer les effets de cette intégration sur les processus cognitifs de l'opérateur, et sur son efficience ;
  - Le deuxième axe est l'opposé du précédent. Il se situe lui aussi dans la perspective du contrôle aérien (pseudo-pilotage ou conditions réelles). Nous faisons l'hypothèse que, passé un certain taux, les occurrences de disfluences pourraient être le signe d'un problème d'ordre cognitif (par exemple, charge cognitive trop importante, ou fatigue). Cette hypothèse est basée sur le fait que, dans certaines situations dites « dégradées » (handicap, environnement ...), les comportements peuvent être identiques (cf. notamment Nespoulous (1996) et Nespoulous *et al.* (2004)). Or, on sait que des disfluences trop nombreuses ou de portée trop grande indiquent généralement un problème pathologique (cf. Blanche-Benveniste *op. cit* ou Rossi *et al. op. cit.*). Il est par conséquent vraisemblable de postuler que des disfluences trop fréquentes, en l'absence de pathologie, sont provoquées par une « dégradation » de la situation du locuteur. Si cette hypothèse s'avérait fondée, reconnaître automatiquement les disfluences permettrait de déterminer, par exemple le moment où un contrôleur doit se reposer<sup>126</sup>.
- Linguistique et TALN : nous avons montré dans les chapitres 3 et 4 les principales implications linguistiques que suggèrent la distribution et le mode de manifestation des disfluences dans notre corpus. Mais ce faisant d'autres questions se posent à nous. En particulier, ces observations sont-elles généralisables à tout type de langage contraint ? Si, comme nous le pensons, la réponse à cette question est affirmative, une recherche intéressante à mener serait de déterminer à partir de quel seuil de « contrainte » du langage et de la

---

<sup>126</sup> A partir d'un seuil à fixer après étude approfondie.

## **Conclusion**

---

tâche les manifestations de disfluences changent par rapport au langage « tout venant ».

Enfin d'une manière plus globale, il sera fructueux d'établir des « ponts » entre les travaux sur les disfluences et ceux sur les erreurs à l'écrit, que nous avons évoqués dans le premier chapitre. La modalité est certes différente. Mais dans les deux cas c'est un humain qui parle/écrit : il y a nécessairement des points communs quant aux stratégies face aux erreurs, aux problèmes provoquées par une situation dégradée, etc. Ces points communs doivent pouvoir se traduire en termes de méthodologie d'analyse et de traitement.

## Bibliographie

- Abney (1991) Abney S. "Parsing by chunks". In R. Berwick, S. Abney, and C. Tenny (Eds.), *Principle based parsing*, Kluwer Academic, 1991.
- Adda-Decker *et al.* (2004) Adda-Decker M., Habert B., Barras C., Adda G., Boula De Mareuil P., Paroubek P.. « Une étude des disfluences pour la transcription automatique de la parole spontanée et l'amélioration des modèles de langage », *JEP'04*. Fez, 2004.
- Antoine *et al.* (2002) Antoine J.-Y., Bousquet C., Goulian J., Zakaria Kurdi M., Rosset S., Vigouroux N., Villaneau J., "Predictive and objective evaluation of speech understanding: the 'challenge' evaluation campaign of the I3 workgroup of the French CNRS", *Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Espagne, Vol. II, p. 529-536, 2002.
- Antoine *et al.* (2003) Antoine J.-Y., Goulian J., Villaneau J.. « Quand le tal robuste s'attaque au langage parlé : analyse incrémentale pour la compréhension de la parole spontanée ». *Actes de TALN 2003*, Batz-sur-Mer, France, p. 25-34, juin 2003.
- (Arrêté J.O. 2000) *Arrêté du 27 juin 2000 relatif aux procédures de radiotéléphonie à l'usage de la circulation aérienne générale*, J.O n° 171 du 26 juillet 2000, p. 11501.
- Athènes *et al.* (2002) Athènes S., Averty P., Puechmorel S., Delahaye D., Collet C. "ATC complexity and controller workload: trying to bridge the gap", *HCI-Aero 2002 International Conference on Human-Computer Interaction in Aeronautics* (p. 56-60) Cambridge, USA aai editor, 2002.
- Austin (1962) Austin J. L. *How to do Things with Words*, Oxford University Press, 1962.
- Averty *et al.* (2002) Averty P., Athènes S., Collet C., Dittmar A., "Evaluating a new index of mental workload in real

- control situation using psychophysiological measures”, *21st Digital Avionics Systems Conference proceedings DASC2002 (7a4p1-13)*, USA October 2002.
- Averty *et al.* (2004) Averty P., Collet C., Dittmar A., Athenes S., Vernet-Maury E. “Mental workload in air traffic control: an index constructed from field tests”. *Aviation, Space, and Environmental Medicine*, vol. 75, iss. 4, p. 333-341(9), April 2004.
- Barras *et al.* (2000) Barras C., Geoffrois E., Wu Z., Liberman M., “Transcriber: development and use of a tool for assisting speech corpora production”, *Speech Communication special issue on Speech Annotation and Corpus Tools*, Vol. 33, No 1-2, 2000.
- Bear *et al.* (1992) Bear J., Dowding J., Shriberg E. “Integrating Multiple Knowledge Sources For Detection And Correction Of Repairs In Human-Computer Dialog”, *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, p.56-63, 28 juin-02 juillet, 1992.
- Bear *et al.* (1993) Bear J., Dowding J., Shriberg E. *A System for Labeling Self-Repairs in Speech*. Technical Note 522, Stanford Research International, février 1993.
- Bernsen *et al.* (2001) Bernsen N. O., Dybkjær L. “Exploring natural interaction in the car”, *CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation*, Verona, Italy, December 2001.
- Blanche-Benveniste (1997) Blanche-Benveniste C *Approches de la langue parlée en français*, Ophrys, Paris, 1997.
- Blanche-Benveniste (2003) Blanche-Benveniste C. « La naissance des syntagmes dans les hésitations et répétitions du parler » In J.L. Araoui (Éd.), *Le sens et la mesure. Hommages à Benoît de Cornulier* (p. 40-55), Éditions Honoré Champion, Paris, 2003.
- Boissière *et al.* (2007) Boissière P., Bouraoui J.-L., Vella F., Lagarrigue A., Mojahid M., Vigouroux N., Nespoulous J.-L.. « Méthodologie d'annotation des erreurs en production écrite. Principes et résultats préliminaires », *Traitement Automatique des Langues Naturelles (TALN 2007)*, Toulouse, p. 529-538, juin 2007.

- Bouraoui *et al.* (2003a) Bouraoui J.-L., Vigouroux N., Bothorel G. (2003) "Transcription and annotation of an apprenticeship corpus: application to the correction and self-correction strategies", *ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, p. 35-40, Chateau-d'Oex-Vaud, Suisse, 28-31 august 2003. ISCA, Stockholm.
- Bouraoui *et al.* (2003b) Bouraoui J.-L., Vigouroux N., « Transcription et annotation d'un corpus d'oral spontané avec le logiciel Transcriber © : illustration sur un corpus de formation. », *INTO'01 (Intonation, Notation and Transcription for Oral structures 01: intonation software)*, Rouen, 22 mai 24 mai 2003.
- Bouraoui *et al.* (2005a) Bouraoui J.-L., Vigouroux N., « Analyse des erreurs de performance et des stratégies correctives dans le dialogue oral spontané : apports à l'étude des pathologies du langage », *Revue Parole*, V. 2004-29-30, p. 121-152, 2005.
- Bouraoui *et al.* (2005b) Bouraoui J.-L., Vigouroux N., "Speech errors management in air traffic control communications: a detailed study". *13th Int. Symposium on Aviation Psychology*, Oklahoma City, 18/04/2005-21/04/2005, avril 2005.
- Bouraoui *et al.* (2005c) Bouraoui J.-L., Vigouroux N., "Disfluencies phenomena in an apprenticeship corpus", *Disfluency in Spontaneous Speech 2005* (DISS 05), Aix-en-Provence, 10/09/2005-12/09/2005, Univ. de Provence, p. 33-37, septembre / september 2005.
- Bouraoui *et al.* (2006a) Bouraoui J.-L., Vigouroux N., « Etude de disfluences dans un corpus linguistiquement contraint », *Journées d'Etudes sur la Parole (JEP 2006)*, Dinard, 12/06/2006-16/06/2006, IRISA, p. 429-432, juin 2006.
- Bouraoui *et al.* (2006b) Bouraoui J.-L., Ehrette Th., Grisvard O. "Spoken Speech Problems Modelling for ATC Dialogue Tracking", *International Conference on Human-Computer Interaction in Aeronautics (HCI'Aero 2006)*, Seattle, USA, 20/09/2006-22/09/2006, F. Reuzeau, K. Corker, G. Boy (Eds.), Cépaduès Editions, p. 148-152, 2006.
- Bouraoui *et al.* (2007) Bouraoui J.-L., Boissière P., Vella F., Lagarrigue A., Mojahid M., Laur D., Vigouroux N., Nespoulous J.-



- L., *Prolégomènes à l'étude des erreurs en production écrite - Propositions en vue de la mise au point d'une grille d'analyse* Rapport de recherche, IRIT/RR-2007-7-FR, IRIT, Toulouse, avril 2007.
- Bousquet (2002) Bousquet-Vernhettes C. *Compréhension robuste de la parole spontanée dans le dialogue oral homme-machine – Décodage conceptuel stochastique*, thèse de doctorat, Université de Paul Sabatier, Toulouse, 2002.
- Bousquet *et al.* (2003) Bousquet-Vernhettes C., Privat R., Vigouroux N. "Error handling in spoken dialogue systems: toward corrective dialogue", *ISCA workshop on Error handling in dialogue systems*, 2003.
- Bousquet *et al.* (2004) Bousquet-Vernhettes C. « Etudes des situations problématiques dans le dialogue oral homme-machine », *XXVe Journées d'Etude sur la Parole (JEP'2004)*, Fès, Maroc, 19/04/2004-21/04/2004, p. 93-96, avril 2004.
- Bove *et al.* (2006) Bove R., Chardenon C., Véronis J., « Prise en compte des disfluences dans un système d'analyse syntaxique automatique de l'oral », *Actes de la Conférence Traitement Automatique des Langues (TALN 2006)*, vol. 1, p. 103-111, Louvain, Belgique, 2006.
- Branigan *et al.* (1999) Branigan H., Lickley R.J., McKelvie D., "Non-linguistic influences on rates of disfluency in spontaneous speech", *Proceedings of the 14th International Congress of Phonetic Sciences*, p. 387-390, San Francisco, 1-7 août 1999.
- Bressolle (2000) Bressolle M.-C., *Modalités et stratégies de construction d'un environnement cognitif commun. Le cas des contrôleurs de la navigation aérienne*, Thèse de doctorat en ergonomie, Université Toulouse le Mirail, 2000.
- Buisson *et al.* (2002) Buisson M., Bustico A., Chatty S., Colin F-R., Jestin Y., Maury S., Mertz Ch., Truillet Ph. « Ivy : un bus logiciel au service du développement de prototypes de systèmes interactifs », *Proceedings of the 14<sup>th</sup> French-speaking conference on HCI, Poitiers*, p.223-226, 2002.
- Barras *et al.* (2000) Barras C., Geoffrois E., Wu Z., Liberman M. (2000). "Transcriber: development and use of a tool for

- assisting speech corpora production”, *Speech Communication special issue on Speech Annotation and Corpus Tools*, Vol 33, No 1-2, 2000.
- Calliope (1989) Calliope (nom collectif), *La parole et son traitement automatique*, Paris Masson, 1989.
- Candéa (2004) Candéa M. *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané*. Thèse d'État, Université Paris III (Sorbonne Nouvelle), 2000.
- Charniak *et al.* (2001) Charniak E., Johnson M. “Edit Detection and Parsing for Transcribed Speech. ”, *Proceedings of NAACL'01*, p. 118-126, 2001.
- Catach *et al.* (1984) Catach, N., Jejcic, F. et équipe HESO, *Les listes orthographiques de base du français (LOB). Les mots les plus fréquents et leurs formes fléchies les plus fréquentes*, Nathan, Paris, 1984.
- Carbonell *et al.* (1983) Carbonell, J. G., Hayes, P.J., (1984), “Recovery strategies for parsing extragrammatical language”, *American Journal of Computational linguistics*, 9(3-4), p. 123-146, 1983.
- Chatty *et al.* (2002) Chatty S., Jestin Y., Maury S., “The Ivy architecture and protocol by Specifications for those who want to understand the internals of Ivy or create new implementations”, *Rapport interne du CENA*, NT02-817, 2002.
- Chomsky (1956) Chomsky, N. “Three models for the description of language”. *IRE Transactions on Information Theory* (2): p.113-124, 1956.
- Cornuéjols *et al.* (2003) Cornuéjols A., Miclet L., Kodratoff Y., *Apprentissage artificiel -Concepts et algorithmes*, 2ème édition, Eyrolles, Paris, 2003.
- Coullon *et al.* (2000) Coullon I., Graglia L. « Spécifications de la base de données pour l'analyse des communications VHF en route », rapport interne du CENA, 2000.
- Coullon *et al.* (2001) Coullon I., Graglia L., Kahn J., Pavet D. « Définition détaillée du document type (DTD) pour le codage sous XML des communications VHF en route – VOCALISE Trafic CRNA / France 2000 », rapport interne du CENA, 2001.

- Davison (2003) Davison J. "When Language Becomes a Barrier Instead of a Bridge: Communication Failures Between Pilots and Air Traffic Controllers", *12th International Symposium on Aviation Psychology*, Dayton, April 14-17, 2003.
- De Calmès *et al.* (1998) De Calmès M., Pérennou.G., "BDLEX : a Lexicon for Spoken and Written French" *1st International Conference on Language Resources & Evaluation*, Grenade, ELRA, Paris, p. 1129-1136, 28-30 mai 1998.
- Dehais (2004) Dehais F. *Modélisation des conflits dans l'activité de pilotage*, thèse de doctorat, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, 2004.
- Deherly *et al.* (2005) Nicolas Deherly, Alexandre Lemort & Philippe Truillet, *ESCALE – CARE II Innovative, WP3.1: Application definition*, Eurocontrol, Brétigny-sur-Orge, 2005.
- Dourmap *et al.* (2002) Dourmap L., *L'interaction vocale dans l'ATC, la formalisation d'une grammaire contextuelle pour la reconnaissance des indicatifs dans le cadre du projet Voice*, Rapport de DEA Informatique, Université Paul Sabatier, Toulouse, 2002.
- Dourmap *et al.* (2003) Dourmap L., Truillet T., « Interaction vocale dans le contrôle aérien : la comparaison de deux grammaires contextuelles pour la reconnaissance des indicatifs de vol », *Rapport interne du CENA*, NR03-669, 2003.
- Dourmap *et al.* (2004) Dourmap L., Truillet T., "Vocal interaction in air traffic control - The VOICE Project", *Int. Conf. on Human-Computer Interaction in Aeronautics (HCI'Aero 2004)*, Toulouse, 29/09/2004-01/10/2004, ACM Press, septembre 2004.
- Drut *et al.* (2008) Drut N., Pierre C., Schon C. « Saisie d'instructions pour le pseudo-pilotage », *Rapport de chef d'œuvre du Master 2 Pro IHM*, ENAC, Toulouse, mars 2008.
- Falzon (1982) Falzon P. « Les communications verbales en situation de travail – Analyse des restrictions du langage naturel », *Rapport technique de l'INRIA*, Rocquencourt, Décembre 1982. Téléchargeable : <http://www.inria.fr/rrrt/rt-0019.html>

- Falzon (1989) Falzon P. *Ergonomie cognitive du dialogue*, Eds: Presses universitaires de Grenoble, 1989.
- Clark *et al.* (2002) Clark, H. H. & Fox Tree, J. E.. "Using uh and um in spontaneous speech". *Cognition*, 84, p. 73-111, 2002.
- Garrett (1980) Garrett M.F., "Levels of processing in sentence production", B.L. Butterworth, *Language Production, volume 1: Speech and Talk*, New York, Academic Press, 1980.
- Graglia *et al.* (2005) Graglia L., Favennec B., Arnoux C., Pavet D., "Vocalise: assessing the impact of Data-Link technology on the R/T channel", *Digital Avionic Systems'05*, Washington, 30 October-3 November 2005.
- Grice (1975) Grice H.-P. "Logic and conversation". In Cole, P. and Morgan, J. (eds.) *Syntax and semantics*, vol 3. Academic Press, New York, 1975.
- Goulian J. (2002) Goulian J., *Stratégie d'analyse détaillée pour la compréhension automatique robuste de la Parole*, Thèse de Doctorat, Université de Bretagne Sud, Vannes, 2002.
- Guénot (2005) Guénot M.L., « Parsing de l'oral: traiter les disfluences », *TALN 2005*, Tome 1, p. 323-332, Dourdan, France, 6-10 juin 2005.
- Hagmueller *et al.* (2006) Hagmueller M., Rank E., Kubin G., *Evaluation of the Human Voice for Indications of Workload Induced Stress in the Aviation Environment*, Eurocontrol Note No. 18/06, Brétigny-sur-Orge, 2006.
- Harmegnies B. *et al.* (1992) Harmegnies B., Landercy A, "A multivariate approach for the analysis of speech under cognitive stress", *ESCA Workshop "Speech Processing in Adverse Conditions"*, p. 231-234, Nice, 1992.
- Hohenberger *et al.* (2004) Hohenberger A., Waleschkowski E. "Speech errors as evidence for language production processes. A historical journey from Meringer to Leuninger", *International Conference on Linguistic Evidence*, Tübingen, 2004.
- Heeman *et al.* (1994) Heeman P., Allen J., "Detecting and correcting speech repairs", *Proceedings of the 32nd annual meeting on*

- Association for Computational Linguistics*, p. 295 – 302, Las Cruces, New Mexico, 1994.
- Henry (2002) Henry S. « Etude des répétitions en français parlé spontané pour les technologies de la parole », *Actes de la 6ème Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL'02)*, (p. 467-476). Nancy (France), 2002.
- Henry *et al.* (2003) Henry S., Pallaud, B. “Word fragments and repeats in spontaneous spoken French”, *DiSS'03, Disfluency in Spontaneous Speech Workshop* (pp. 77-80). Göteborg (Sweden), 2003.
- Henry *et al.* (2004) Henry S., Campione, E., Véronis J. « Répétitions et pauses (silencieuses et remplies) en français spontané », *XXVèmes Journées d'Etude sur la Parole (JEP'04)*, Fès (Maroc), p. 261-264, 2004.
- Hirschberg *et al.* (2004) Hirschberg J., Litman D., Swerts M., “Identifying user corrections automatically in spoken dialogue systems”, *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, p. 1-8, 2001.
- Hindle (1983) Hindle D., “Deterministic parsing of syntactic non-fluencies”, *Proceedings of the 21st annual meeting on Association for Computational Linguistics* Cambridge, Massachusetts, p.123 - 128, 1983
- Jakobson (1963) Jakobson R « Linguistique et poétique », *Essais de linguistique générale*, Paris, Minuit, p. 209-248 ; 1963.
- Johnson *et al.* (2004a) Johnson M., Charniak E., Lease M.. “An Improved Model For Recognizing Disfluencies in Conversational Speech”, *Rich Transcription 2004 Fall Workshop (RT-04F)*, 2004.
- Johnson *et al.* (2004b) Johnson M., Charniak E., “A Tag-Based Noisy Channel Model of Speech Repairs”, *ACL'04*, p. 33-39, 2004.
- Kerbrat-Oreccioni (1995) Kerbrat-Oreccioni C. *Les interactions verbales*, Armand Colin, Paris, 1995.
- Kremin (1994) Kremin H. « Perturbations lexicales : les troubles de la dénomination », in Seron X, Jeannerod M. (sous la

- direction de), *Neuropsychologie humaine*, Liège, Mardaga, p. 375-389, 1994.
- Kurdi (2003) Kurdi M.- Z. *Contribution à l'analyse du langage oral spontané*, Thèse de doctorat, Université J. Fourier, Grenoble, 18 avril 2003.
- Labelle (2002) Labelle M. « Trente ans de psycholinguistique », *Revue québécoise de linguistique*, Volume 30, numéro 1, Université du Québec, 2002.
- Lacheret *et al.* (1999) Lacheret Dujour, A., Beaugendre F. *La prosodie du français*, CNRS édition, Paris, 1999.
- Lease *et al.* (2006) Lease M., Johnson, M., Charniak E., “Recognizing disfluencies in conversational speech”, *IEEE Transactions on Audio, Speech, and Language Processing*, Volume 14, Issue 5, p. 1566 – 1573, Sept. 2006.
- Lerat (1995) Lerat P. *Les langues spécialisées*. Paris, PUF, 1995.
- Levelt *et al.* (1999) Levelt W. J. M., Roelofs A. Meyer A. S., “A theory of lexical access in speech production”, *Behavioral And Brain Sciences*, 22, p. 1–75, 1999.
- Levenshtein (1966) Levenshtein V.I. “Binary codes capable of correcting deletions, insertions, and reversals”, *Cyber. Contr. Theory*, 10 (8), p. 707-710, 1966.
- Lickley (1994) Lickley R.J., *Detecting Disfluency in Spontaneous Speech*, Ph.D. thesis, University of Edinburgh, 1994, 1994.
- Liu (2003) Liu Y., “Word fragment identification using acoustic-prosodic features in conversational speech”, *HLT-NAACL student research workshop*, p. 37-42, 2003.
- Liu *et al.* (2003) Liu Y., Shriberg E., Stolcke A., Hillard D., Ostendorf M., Harper M., “Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies”, *IEEE Transactions on Audio, Speech, and Language Processing*, V14(5), p. 1526-1540, September 2006.
- Luzzati (2004) Luzzati D., « Le fenêtrage syntaxique : une méthode d'analyse et d'évaluation de l'oral spontané », *MIDL 2004*, 29-30 novembre 2004, Paris.

- Marque *et al.* (1995)      Marque F., Salvi R., « Résultats de l'évaluation de PAROLE à l'ENAC », *Rapport interne du CENA*, N95086, 1995.
- Martinet (1960)      Martinet A., *Éléments de linguistique générale*, Armand Colin, 4<sup>ème</sup> édition, 2003.
- Maugis (1995)      Maugis L., *VOCALISES Speech Interface: Preliminary Studies*, version 1.0, CENA note, N95051, 1995.
- McTait *et al.* (2004)      McTait K., Maynard H., Devillers L., Rosset S., Paroubek P., Mostefa D., Choukri K., Antoine J-Y, Béchet F., Bousquet C., Vigouroux N., Bontron O., Charnay L., Romary Lt, Vergnes M., « Constitution d'un corpus de dialogue oral pour l'évaluation automatique de la compréhension hors -et en- contexte du dialogue », *JEP 2004*, Fès, Maroc, p. 357-360, 19-22 avril 2004.
- Mertz *et al.* (2000)      Mertz C., Chatty S., Vinot J.-L., "The influence of design techniques on user interfaces: the DigiStrips experiment for air traffic control", *HCI-Aero 2000 International Conference on Human-Computer Interaction in Aeronautics*, Toulouse, France, 27-29 September 2000
- Mertz (2003)      Mertz C., "Peripheral awareness offered by interaction techniques in Air Traffic Control interfaces", *CHI 2003 New horizons Conference - Workshop on "Providing elegant peripheral awareness"*, Fort-Lauderdale, 5-10 April 2003.
- Moeschler (1989)      Moeschler J., *Modélisation du dialogue*, Hermès, Paris, 1989.
- Moeschler *et al.* (1994)      Moeschler J., Reboul A. *Dictionnaire encyclopédique de pragmatique*, Seuil, Paris, 1994.
- Minker (1999)      Wolfgang Minker, *Compréhension automatique de la parole spontanée*, L'Harmattan, 1999.
- Nakatani *et al.* (1994)      Nakatani C. H., Hirschberg, J., "A corpus-based study of repair cues in spontaneous speech", *The Journal of the Acoustical Society of America*, Volume 95, Issue 3, p. 1603-1616, Mars 1994.
- Nespoulous (1990)      Nespoulous J.-L. « De la difficulté d'interprétation des manifestations linguistiques de surface », *Linguistique*

- et neuropsycholinguistique : tendances actuelles*, Société de neuropsychologie de langue Française, p. 5-15, 1990.
- Nespoulous (1996) Nespoulous J.-L. « Les stratégies palliatives dans l'aphasie », *Rééducation Orthophonique*, Vol. 34, Décembre 1996, N° 188, p. 423-433, 1996.
- Nespoulous *et al.* (2004a) Nespoulous J.-L., Virbel J., « Apport de l'étude des handicaps langagiers à la connaissance du langage humain », revue *PArole*, Vol. 29/30, Université de Mons-Hainaut, (Belgique), p. 5-42, 2004.
- Nespoulous (2004b) Nespoulous J.-L. « Linguistique, Pathologie du Langage et Cognition : des dysfonctionnements langagiers à la caractérisation de l'architecture fonctionnelle du langage », in C. Fuchs (Ed.) *Les linguistiques cognitives*, Ophrys & Editions de la MSH, « Collection Cogniprismes », 2004.
- O'Shaughnessy (1992) O'Shaughnessy D., "Analysis of false starts in spontaneous speech", *Proceedings of the International Conference on Spoken Language Processing*, pp. 931-934, Banff, 1992.
- Oviatt (1995) Oviatt S. "Predicting Spoken Disfluencies during Human-Computer Interaction". *Computer Speech Language*, Vol 9, No 1, p. 19-36, 1995.
- Pallaud *et al.* (2004) Pallaud B., Henry S. « Amorce de mots et répétitions : des hésitations plus que des erreurs en français parlé », *Le poids des mots, Actes des 7èmes Journées Internationales d'Analyse statistique des Données Textuelles*, PUL, vol 2 : p. 848-858, Louvain, 10-12 mars 2004.
- Perennou (1996) Pérennou G., « Compréhension du dialogue oral – Le rôle du lexique dans l'approche par segments conceptuels », *Lexique et communication parlée - GDR PRC*, p. 169-178, 1996.
- Pieraccini *et al.* (1995) Roberto Pieraccini et Esther Levin, "A Spontaneous-Speech Understanding System for Database Query Applications", *ESCA Workshop on Spoken Dialogue Systems*, Vigso, Danemark, p. 85-88, 1995.
- Pierrel (2000) Pierrel J.-M., *Ingénierie des langues*, Hermès science Europe, 2000.



- Pillon *et al.* (1994) Pillon A., Nespoulous J.-L. « Perturbations syntaxiques dans le langage aphasique », in Seron X, Jeannerod M. (sous la direction de), *Neuropsychologie humaine*, Liège, Mardaga, p. 360-374, 1994.
- Piu *et al.* (2007) Piu M., Bove R., « Annotation des disfluences dans les corpus oraux », *TALN'07, RÉCITAL 2007*, p. 397-406, Toulouse, 5–8 juin 2007.
- Plauché *et al.* (1999) Plauché M. C., Shriberg E., “Data-Driven Subclassification Of Disfluent Repetitions Based On Prosodic”, *Proceedings of the 14th International Congress of Phonetic Sciences*, p. 1513-1516, San Francisco, 1-7 août 1999.
- Privat *et al.* (2001) Privat R., Vigouroux N., Bousquet C., Truillet P., Oriola B., “Speech Technology for Universal Access in Interactive Systems”, *1st International UAHCI 2001 Conference*, 05/08/2001-10/08/2001, Gavriel Salvendy, p. 416-420, New Orléans, août 2001.
- Rasmussen (1983) Rasmussen J., “Skills, rules, knowledge: signals, sign and symbols and other distinctions in human performance models”, *IEEE Transactions: Systems, Man & Cybernetics; SMC-13*, p. 257-267, 1983.
- Reason (1990) Reason J. *Human error*, Eds: Cambridge University Press, 1990.
- Roulet *et al.* (1985) Roulet E., Auchlin A., Moeschler J., Schelling M., Rubattel C. *L'articulation du discours en français contemporain*, Berne, Lang (Collection Sciences pour la communication), 1985.
- Rossi *et al.* (1998) Rossi M., Peter-Defare E. *Les lapsus, ou comment notre fourche a langué*, PUF, Paris, 1998.
- Schäfer (2001) Schäfer D., *Context-Sensitive Speech Recognition in the Air Traffic Control Simulation*, Eurocontrol Experimental Center (EEC) Note n°02/2001, février 2001.
- Scherer (1986) Scherer K. R., *Voice, stress, and emotion, Dynamics of stress*, M. H. Appley, & R. Trumbull, (Eds.), New York: Plenum, pp. 159- 181, 1986.
- Schegloff (1977) Schegloff E., Jefferson G., Sacks H., “The preference for self-correction in the organization of repair in conversation”, *Language* 52 (2), p. 361-382, 1977.

- Searle (1972) Searle J. R. (1972) *Les actes du langage*, Paris, Hermann.
- Seron *et al.* (1994) Seron X., Jeannerod M. (sous la direction de), *Neuropsychologie humaine*, Liège, Mardaga, 1994.
- Shin *et al.* (2002) Shin J., Narayanan S., Gerber L., Kazemzadeh A., Byrd D. “Analysis of user behaviour under error conditions in spoken dialogs”, *Proceedings of the Seventh International Conference on Spoken Language Processing*, Colorado, September, 2002.
- Shriberg *et al.* (1992) Shriberg, E. E., Bear, J. & Dowding, J., “Automatic detection and correction of repairs in human-computer dialog,” *Proceedings of the DARPA Speech and Natural Language Workshop*, p. 419-425, 23-26 février 1992.
- Shriberg (1994) Shriberg E.. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of Berkeley, California, 1994.
- Shriberg *et al.* (2004) Shriberg E., Stolcke A., “Prosody modeling for speech recognition and understanding”, *Mathematical Foundations of Speech and Language Processing* (M. Johnson, S. Khundapur, M. Ostendorf, and R. Rosenfeld, eds.), vol. 138 of *IMA Volumes in Mathematics and its Applications*, p. 105-114, Springer, New-York, 2004.
- Spérandio (1984) Sparandio J. C., *L'ergonomie du travail mental*, Masson, Paris, 1984.
- Stolcke (1996) Stolcke A., Shriberg E., “Statistical language modeling for speech disfluencies”, *Proc. ICASSP*, vol. 1, pp. 405–408, Atlanta, May 1996.
- Traum (1999) Traum D. R., “Computational Models of Grounding in Collaborative Systems”, *Working notes of AAAI Fall Symposium on Psychological Models of Communication*, p. 124-131, November, 1999.
- Truillet *et al.* (2001) Truillet P., Vigouroux N. (2001) « Interaction vocale dans l'ATC : le projet VICTOR », *IHM-HCI'2001*, p. 179-182, Lille, 10-14 septembre 2001.
- Truillet *et al.* (2004) Truillet P., Grisvard O., Goujon B., *SCOPE – CARE II Innovative WP3 – R3 – Model of English Callsigns*, Eurocontrol, Brétigny-sur-Orge, 2004.

- Truillet *et al.* (2005) Truillet P., Bothorel G. (2005) « VOICE : une plateforme pour la formation au contrôle aérien », *IHM 2005*, p. 67-80, Toulouse, 27-30 septembre 2005.
- Valdois *et al.* (1994) Valdois S., Nespoulous J.-L. « Perturbations du traitement phonétique et phonologique du langage », in Seron Xavier, Jeannerod Marc (sous la direction de), *Neuropsychologie humaine*, Liège, Mardaga, 1994.
- Valli *et al.* (1999) Valli A., Véronis J., « Etiquetage grammatical de corpus oraux : problèmes et perspectives », *Revue Française de Linguistique Appliquée*, Vol. IV(2), pp. 113-133, 1999.
- Vigouroux (2001) Vigouroux N., *Cours de DEA Informatique de l'Image et du Langage*, Toulouse, Université Paul Sabatier, 2001.
- Veronis (1991) Veronis J. "Error in natural language dialogue between man and machine", *International journal of man-machine studies*, Academic Press, vol. 35, no2, p. 187-217, 1991.
- Volpe *et al.* (2002) Volpe R., Pavet D. *Vocalise : Une analyse du canal vocal pilotes-contrôleurs dans la perspective d'un environnement data-link - Trafic Approche / France*; rapport CENA, 2002.
- Volpe *et al.* (2004) Volpe R., Pavet D. *Vocalise : Une analyse du canal vocal pilotes-contrôleurs dans la perspective d'un environnement data-link - Trafic Sol/Prévol* ; rapport CENA, 2004.
- Wall *et al.* (2001) Wall L., Christiansen T., Orwant J., *Programmation en Perl*, O'Reilly, 2001.

## Annexe 1 : Description de l'algorithme principal

Dans cette annexe, nous présentons l'algorithme gérant les disfluences. Cet algorithme est décomposé en plusieurs fonctions. Nous ne décrivons que celles participant au traitement des disfluences, et non les autres fonctions relatives à la gestion de fichiers, la connexion à IVY, la transformation des chaînes de caractères, etc. Nous présentons cependant rapidement, avant l'algorithme, quelques fonctions non détaillées dans celui-ci mais dont le rôle est malgré tout important à connaître.

Selon les conventions en usage, les énoncés écrits en caractères droit correspondent au code ; ceux en *italique*, et précédés de deux slash (« // ») ou compris entre les caractères « /\*\*\* » et « \*\*\*/ », sont des commentaires. Pour la désignation des fonctions, des variables et des tableaux, nous adoptons les conventions du langage *PERL* : respectivement `&nom_fonction(nom_paramètre1, nom_paramètre2,...)`, `$nom_variable`, `@nom_tableau`. Enfin, les symboles « = = » et « = » désignent respectivement la comparaison et l'instanciation.

Pour des raisons de clarté, le pseudo code présenté dans cette annexe ne correspond pas exactement à son équivalent implémenté.

### Rôle des fonctions non détaillées :

- *&Est\_Editeur* : reçoit en entrée deux positions dans un tableau. Vérifie si chaque élément compris entre les deux positions de ce tableau est un éditeur (hésitation, mot « correction », mot « sorry », etc.) ; retourne 1 si le résultat est positif, 0 autrement ;
- *&Est\_Repetition\_Dans\_Niveau*, *&Est\_Repetition\_Dans\_Frequence* : fonctionnent sur le même principe que *&Est\_Repetition\_Dans\_Indicatif* (détaillé dans l'annexe). Autrement dit, elles permettent de déterminer si une répétition observée est provoquée par la phraséologie (par exemple l'expression d'un niveau : « 1 1 0 ») et est donc non disfluente ;
- *&Tag\_Disfluence* : ajoute une balise à la disfluence, en fonction de son type. Les balises existantes sont « *Repetition* », « *Element\_Repete* », « *Element\_corrige* », « *Correction* » ;

## Annexe 1 : Description de l'algorithme principal

---

- *&Construit\_Sortie* : crée la chaîne de caractères qui sera renvoyée sur IVY. Son rôle consiste essentiellement à reconstruire l'énoncé traité, en prenant en compte les ajouts éventuels faits par la fonction *&Tag\_Disfluence* ;
- *&Parse\_enonce* : accomplit un passage de l'énoncé, depuis le début de celui-ci jusqu'à la position qui lui est indiqué en paramètre. Pour des raisons techniques, le moteur de passage utilisé par cette fonction est différent de celui, utilisé par le module *Sra*, qui a abouti à la décomposition de l'énoncé en paires attribut=valeur.

```
#-----  
# Objet : Fonction principale du module.  
#   Prend en entrée une ligne, la décompose, l'analyse  
#   et retourne en sortie une ligne étiquetée.  
#-----
```

Traite\_Ligne {

*// Décomposition des éléments de l'énoncé*

    Instanciation du tableau @classes avec les classes de mot de l'énoncé *//Une classe par cellule*

    Instanciation du tableau @mots avec les mots de l'énoncé *//Un mot par cellule*

    Instanciation de \$nombre avec le nombre total de mots et de classes de mots de l'énoncé

*// Définition de la fenêtre coulissante*

    Pour \$debut\_fenetre = 0 à \$debut\_fenetre < \$nombre - 1 {

        \$fin\_fenetre = \$i + 10 - 1;

        Si (\$fin\_fenetre >= \$nombre) Alors { \$fin\_fenetre = \$nombre - 1; }

        @Fenetre = ensemble des positions entre \$debut\_fenetre et \$fin\_fenetre

*// Recherche de disfluences*

        Pour chaque mot de @Fenetre Faire

            Comparer ce mot avec chaque autre mot de @Fenetre

                Si les deux éléments sont identiques Alors

                    Récupérer les positions de l'élément répété (\$pos1) et celle de la répétition (\$pos2)

                    Appeler la fonction *&Traite\_Repetition\_Lexicale*

                Sinon

                    Pour chaque classe de mots de @Fenetre Faire

                        Comparer cette classe de mots avec chaque autre classe de mots de @Fenetre

                            Si les deux éléments sont identiques Alors

                                Récupérer les positions de l'élément répété (\$pos1) et celle de la répétition (\$pos2)

Appeler la fonction &Traite\_AutoCorrection

```

Sinon //suite du parsage
    $debut_fenetre = $debut_fenetre+1
    $fin_fenetre = $fin_fenetre+1
}
}

-----

#-----
# Objet : Traite le cas des répétitions lexicales (deux mots identiques).
# Vérifie que les mots spécifiés correspondent bien à une répétition,
# et si tel est le cas, les étiquette en conséquence.
# Paramètres :
# - @classes, @mots : Données de la phrase.
# - pos1, pos2 : Position des deux mots répétés dans la phrase.
# Résultat : Retourne vrai (1) si la recherche a abouti, faux (0) sinon.
#-----

Traite_Repetition_Lexicale (@classes, @mots, $pos1, $pos2) {

    // Test des cas particuliers de répétition 'acceptés' (qui ne sont pas des disfluences).
    Appel de la fonction &Est_Cas_Particulier_De_Repetition (@classes, @mots, $pos1,
                                                                $pos2)

    Si (&Est_Cas_Particulier_De_Repetition == 1) {
        // C'est un cas particulier, on sort.
        retourne (0);
    }
    Sinon
    {

        // Teste si la répétition est disfluente :
        // - si la répétition suit immédiatement l'élément répété
        // - s'il n'y a que des hésitations ou équivalent entre la répétition et
        // l'élément répété

        Appel à la fonction &Est_Editeur (@classes, $pos1 + 1, $pos2 - 1)
        Si ($pos2 == $pos1 + 1) OU
        Si (&Est_Editeur == 1) Alors {
            // La répétition est disfluente : étiquetage
            Appel de la fonction &Tag_Disfluence ;
        }
    }
}
}

```

## Annexe 1 : Description de l'algorithme principal

---

```
#-----  
# Objet :  
# Le but des deux fonctions suivantes est de déterminer  
# si l'on a affaire ou non à une répétition provoquée par la phraséologie  
# (indicatifs, niveaux, fréquence).  
# Une telle répétition ne doit pas être étiquetée 'disfluente'.  
# Paramètres :  
# - classes, mots : Données de la phrase.  
# - pos1, pos2 : Position des deux mots répétés dans la phrase.  
# Résultat : Retourne vrai (1) si la recherche a abouti, faux (0) sinon.  
#-----
```

```
Est_Cas_Particulier_De_Repetition (@classes, @mots, $pos1, $pos2) {  
  
    Appel de la fonction &Est_Repetition_Dans_Indicatif (@classes, @mots, $pos1,  
                                                         $pos2)  
  
    // 1) Répétition dans indicatif. Ex: vol 0 0 7.  
  
    Si (&Est_Repetition_Dans_Indicatif == 1) {retourne (1); }  
  
    Sinon  
  
        Appel de la fonction &Est_Repetition_Dans_Niveau (@classes, @mots, $pos1,  
                                                         $pos2)  
  
        // 2) Répétition dans niveau. Ex: niveau 1 1 0.  
  
        Si (&Est_Repetition_Dans_Niveau == 1) { retourner (1); }  
  
        Sinon  
  
            Appel de la fonction & Est_Repetition_Dans_Frequence  
                                (@classes, @mots, $pos1,  
                                $pos2)  
  
            // 3) Répétition dans fréquence. Ex: fréquence 2 0 0.  
  
            Si (&Est_Repetition_Dans_Frequence == 1) {retourne (1);}   
  
            Sinon  
  
                //La répétition n'est pas un cas particulier  
  
                {retourne (0);}   
  
    }  
}
```

```
Est_Repetition_Dans_Indicatif (@classes, @mots, $pos1, $pos2) {  
  
    /**le tableau @items_phraseo comporte l'ensemble des classes de mots impliquées  
        dans la composition des niveaux, fréquences, et indicatifs***/  
  
    @items_phraseo = ( 'Alphabet', 'Chiffre', 'Dizaine',  
                       'Compagnie_Domestique', 'Compagnie_Internationale',  
                       'Mot_Decimale',  
                       );  
  
    /** Un indicatif n'étant formé que de mots de classe alphabet, chiffre, dizaine, et  
        Compagnie (Domestique ou Privée), on vérifie d'abord que les deux mots concernés  
        appartiennent à ces classes.***/  
  
    Si (@classes[$pos1] appartient à @items_phraseo) ET  
  
    Si (@classes[$pos2] appartient à @items_phraseo) Alors  
  
        // On récupère ensuite la séquence qui forme l'indicateur  
  
        //a) Récupération de la limite inférieure 'imin'.  
  
        $imin = $pos1;  
        Tant que ($imin > 0) ET (@classes[$imin] appartient à @items_phraseo) Alors  
        {  
            $imin=$imin -1;  
        }  
  
        //b) Récupération de la limite supérieure 'imax'.  
        $imax = $pos1;  
        Tant que ($imin > 0) ET (@classes[$imin] appartient à @items_phraseo) Alors  
        {  
            $imax=$imax +1;  
        }  
  
        //La séquence va de imin à imax inclus.  
        // On vérifie que pos2 appartient bien à cette séquence  
  
        Si ($pos2 > $imax) {  
            retourne (0);  
        }  
  
        // Enfin, on vérifie que la séquence formée correspond à un indicatif existant.  
  
        $test_indicatif = ensemble des mots situés entre $imin et $imax  
        Si $test_indicatif appartient à @liste_indicatifs Alors  
        {retourne 1 ;} //il s'agit d'un indicatif ; la répétition n'est pas disfluente
```



## Annexe 1 : Description de l'algorithme principal

---

Sinon

{retourne 0 ;} //ce n'est pas un indicatif : la répétition est disfluente

#-----

# Objet : vise à déterminer si une la disfluence est une auto-correction

#-----

# 1) On teste d'abord les trois schémas d'auto-correction les plus fréquents : R1 / R1, R1 M1 / R1 M1, et M1 R1 / M1 R1

# - @classes, @mots : Données de la phrase.

# - pos1 : Position du mot répété dans la phrase.

# - pos2 : Position de la répétition dans la phrase.

# Résultat : Retourne vrai (1) si la recherche a abouti, faux (0) sinon.

#-----

Traite\_AutoCorrection (@classes, @mots, \$pos1, \$pos2) {

    // Test des cas particuliers de répétition 'acceptés' (qui ne sont pas des disfluences).

    Appel de la fonction & Est\_Cas\_Particulier\_De\_Repetition(@classes, @mots, \$pos1, \$pos2)

    Si (&Est\_Cas\_Particulier\_De\_Repetition == 1) Alors {retourne 0;} // C'est un cas particulier, on sort.

    // On teste d'abord les trois schémas d'auto-correction les plus fréquents : R1 / R1, R1 M1 / R1 M1, et M1 R1 / M1 R1

    Si ((&Est\_Motif\_R1R1(@classes, \$pos1, \$pos2) == 1)

        OU

    Si (&Est\_Motif\_R1M1R1M1(@classes, \$ref\_mots, \$pos1, \$pos2) == 1)

        OU

    Si (&Est\_Motif\_M1R1M1R1 (@classes, \$ref\_mots, \$pos1, \$pos2) == 1)

) Alors {Appel à la fonction &Tag\_Disfluence;}

    Sinon {Appel à la fonction &Analyse\_par\_regle}

}

#-----

# Objet : Recherche le motif R1 / R1 dans la séquence donnée.

# Paramètres :

# - classes : Données de la phrase.

# - pos1, pos2 : Position des mots qui représentent les R1 du motif.

# Résultat : La fonction retourne vrai (1) si le motif est trouvé, faux (0) sinon.

#-----

Est\_Motif\_R1R1 (@classes, @mots, \$pos1, \$pos2) {

```

Si ($pos2 == $pos1 + 1) {
    // Les 2 mots se suivent, le schéma est respecté.
    retourne (1);
}

Sinon
    Si (&Est_Editeur (@classes, $pos1 + 1, $pos2 - 1) == 1) {
        /** Entre les 2 mots, il n'y a que des hésitations ou des corrections, le schéma est respecté. */
        retourne (1);
    }

    Sinon
        // Schéma non trouvé.
        {retourne (0)};
}

#-----
# Objet : Recherche le motif R1 M1 / R1 M1 dans la séquence donnée.
# Paramètres :
# - classes, mots : Données de la phrase.
# - pos1, pos2 : Position des mots qui représentent les R1 du motif.
# Résultat : La fonction retourne vrai (1) si le motif est trouvé, faux (0) sinon.
#-----

Est_Motif_R1M1R1M1 (@classes, @mots, $pos1, $pos2) {

    // 1) Les mots qui suivent doivent être identiques (les 'M1').
    Si ($pos2 + 1 > $nombre) { retourne(0); } // pos2 est le dernier mot, il ne peut y avoir de M1.

    Si (@mots[$pos1 + 1] ≠ @mots[$pos2 + 1]) { retourne(0); }

    // 2) Les séquences doivent se suivre
    Si ($pos2 == $pos1 + 2) {
        // Les 2 séquences se suivent, le schéma est respecté.
        retourne (1);
    } Sinon
        Si (&Est_Editeur (@classes, $pos1 + 1, $pos2 - 1)) {
            // Entre les 2 séquences, il n'y a que des éditeurs ou des corrections, le schéma est respecté.
            retourne (1);
        }

    // Schéma non trouvé.
    retourne (0);
}

```

```
#-----
# Objet : Recherche le motif M1 R1/ M1 R1 dans la séquence donnée.
# Paramètres :
# - classes, mots : Données de la phrase.
# - pos1, pos2 : Position des mots qui représentent les R1 du motif.
# Résultat : La fonction retourne vrai (1) si le motif est trouvé, faux (0) sinon.
#-----
Est_Motif_M1R1M1R1 (@classes, @mots, $pos1, $pos2) {

    // 1) Les classes de mots qui suivent doivent être identiques (les 'R1').

    Si ($pos2 + 1 > $nombre) Alors
        {retourne (0);} // pos2 est le dernier mot, il ne peut y avoir de M1.

        Sinon
            Si (@classes[$pos1 + 1] = @classes[$pos2 + 1]) ET
                ($pos2 == $pos1 + 2)) {
                retourne (1); // Les 2 séquences se suivent, le schéma est
                    respecté.
            }

            Sinon
                Appel de la fonction &Est_Editeur (@classes, $pos1 + 1, $pos2
                    - 1)
                Si (&Est_Editeur (@classes, $pos1 + 1, $pos2 - 1) == 1) {
                // Entre les 2 séquences, il n'y a que des éditeurs ou des
                    corrections, le schéma est respecté.
                retourne (1);
                }

        // Schéma non trouvé.
        retourne (0);

    }

#-----
# Objet : Si aucun schéma n'a été trouvé jusqu'à ce point du programme,
# cela veut dire que l'on a affaire à un patron complexe. Son analyse
# nécessite notamment un passage de l'énoncé en cours
# Paramètres :
# - classes, mots : Données de la phrase.
# - pos1, pos2 : Position des mots qui représentent les R1 du motif.
# Résultat : La fonction retourne, le cas échéant, les positions des
# additions et/ou omissions et leur statut
#-----

Analyse_par_regle (@classes, @mots, $pos1, $pos2) {
```

*// Lancement d'une analyse par règles à partir du début de l'énoncé jusqu'à \$pos2*

Appel de la fonction &Parse\_enonce(@mots, \$pos1, \$pos2)

*/\*\*A partir du moment où le passage concerne le contenu de la fenêtre où se trouve une éventuelle disfluence, on cherche La présence d'une interruption de règle indique la présence d'un point d'édition. Il faut maintenant faire la comparaison entre les différentes cases du tableau qui se trouve avant le point d'édition et celles qui se trouvent après \*\*\*/*

Pour @mots[\$pos1] à @mots[\$pos2]

    Si on ne peut accéder à la règle suivante de la grammaire OU

    Si on ne trouve pas de nœud terminal

        Alors

            Arrêt de la fonction &Parse\_enonce

            \$point\_interruption = = @mots[\$pos1+n] *//n est la position dans @mots de l'élément sur lequel la passage s'est arrêté inopinément*

*/\*\*La position du point d'édition est pour l'instant confondue avec celle de l'élément au niveau duquel la règle a été interrompue \*\*\*/*

\$Pos\_point\_ed = = \$point\_interruption

*/\*\* On partage le tableau @mots en 2 tableaux. Les frontières du premier sont délimitées à gauche par la position de l'élément faisant l'objet de la répétition, à droite par la position du point d'édition. \*\*\*/*

Instanciation du tableau @Fenetre\_gauche par la plage comprise entre @mots[\$pos1] et @mots[\$Pos\_point\_ed]

Instanciation du tableau @Fenetre\_droite par la plage comprise entre @mots[\$Pos\_point\_ed+1] et @mots[\$pos2]

*/\*\* On cherche s'il y a une deuxième répétition de mots après pos2. Le but est de détecter d'éventuels schémas de type M1 M2 Mn(...) | (...) M1 M2 Mn ; Le cas échéant, les positions des éléments répétés déjà détectés sont maintenues en mémoire\*\*\*/*

Pour chaque élément de la plage comprise entre @Fenetre\_gauche[\$pos1+1] et @Fenetre\_gauche[\$Pos\_point\_ed] faire {

    Comparer l'élément courant avec @mots[\$Pos2+1]

    Si un élément identique est trouvé, alors

        \$Pos1 = \$pos1+n *//Changement de la borne gauche de la fenêtre (n est la position de l'élément courant)*

        \$Pos2 = \$pos2+n *//Changement de la borne droite de la fenêtre (n est la position de l'élément courant)*

    }

## Annexe 1 : Description de l'algorithme principal

---

*/\*\* On cherche les cas d'addition ou d'omission (les déplacements étant considérés comme des substitutions, et ces dernières ayant déjà été traitées en amont) \*\*/*

```
Pour chaque élément de la plage comprise entre @Fenetre_gauche[$pos1+1] à
@Fenetre_gauche[$Pos_point_ed], Faire {
    Chercher cet élément dans @Fenetre_droite
    Si cet élément est absent de @Fenetre_droite ET
    Si (&Est_Editeur == 0) // Si l'élément courant est un éditeur, on ne le prend
        pas en compte dans le recensement des omissions
        Alors
        Le marquer comme une omission
    }

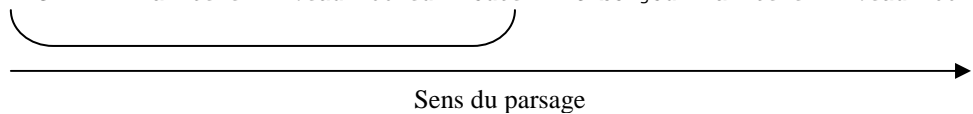
Pour chaque élément de la plage comprise entre @Fenetre_droite[$Pos_point_ed+1] à
@Fenetre_droite[$Pos2], Faire {
    Chercher cet élément dans @Fenetre_gauche
    Si cet élément est absent de @Fenetre_gauche ET
    Si (&Est_Editeur == 0) // Si l'élément courant est un éditeur, on ne le prend pas en
        compte dans le recensement des additions
        Alors
        Le marquer comme une addition
    }
}
```

## Annexe 2 : Exemple d'application de l'algorithme

❶

Initialisation du passage avec la fenêtre.

F G A L Z maintenez niveau 100 euh route ENAC bonjour maintenez niveau 100 route Balon




Sens du passage

❷

(...) On passe les étapes pendant lesquelles la fenêtre ne rencontre pas de disfluences potentielles.

F G A L Z **maintenez** niveau 100 euh route ENAC bonjour **maintenez** niveau 100 route Balon

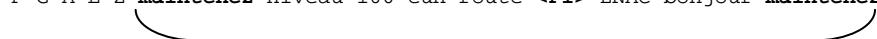


La fenêtre détecte deux similarités lexicales. Le passage s'arrête, et la série de tests commence.

❸

(...) La répétition détectée n'est pas due à la phraséologie, et les trois patrons les plus fréquents ne correspondent pas → appel à la fonction *&Analyse\_par\_regle*. Le point d'*Interruption* est placé à la position où le passage par règle s'arrête inopinément.

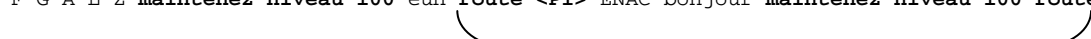
F G A L Z **maintenez** niveau 100 euh route <PI> ENAC bonjour **maintenez** niveau 100 route Balon



❹

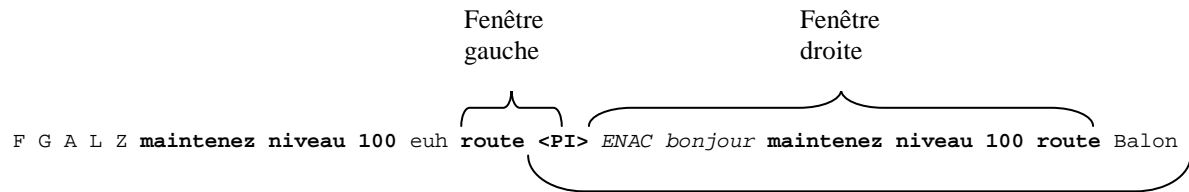
(...) La recherche de patrons de type *M1 M2 Mn(...)* / *(...) M1 M2 Mn* a permis d'établir que d'autres éléments lexicaux sont répétés de part et d'autre du point d'*Interruption*. Les positions sont mises à jour (et gardées en mémoire pour les répétitions déjà détectées).

F G A L Z **maintenez niveau 100 euh route** <PI> ENAC bonjour **maintenez niveau 100 route** Balon



5

(...) Une comparaison entre la « fenêtre gauche » et la « fenêtre droite » révèle deux éléments additionnels au niveau de la « fenêtre droite ». Ils sont étiquetés comme tels (en italique sur l'exemple). Le passage peut reprendre jusqu'à la fin de l'énoncé.



## Annexe 3 : Exemples de transcriptions

Cette annexe présente la transcription de deux dialogues du corpus A, tels que nous les avons effectuées sous *Transcriber*. Nous présentons cette transcription sous deux formats différents :

- Le premier est une représentation conçue pour la consultation « manuelle » de la transcription. Nous l'avons obtenue en exportant au format *HTML*<sup>127</sup> les transcriptions sélectionnées. Grâce à ce format, les différents éléments (locuteurs, tours de parole, phénomènes, etc.) de la transcription sont clairement identifiés au moyen de leur mise en forme.

Voici les différents composants qui sont visibles :

- Report-dialogue : indique le début d'une section ;
  - pil h ang inde, pil h fr inde : désigne deux pilotes qui parlent respectivement en anglais et en français ;
  - ctrl h ang 1, ctrl h fr 1 : désigne deux contrôleurs qui parlent respectivement en anglais et en français ;
  - (no speaker) : indique une période de silence située, dans une section, entre deux tours de parole.
- Le second est le format natif (*.trs*) dans lequel les transcriptions sont enregistrées. Comme nous le disons dans le chapitre 2, ce format est compatible *XML*. Il repose donc sur l'utilisation de balises ouvrantes et fermantes qui permettent de marquer et identifier les chaque élément de la transcription. Aucune mise en forme n'est effectuée, mais les balises permettent une identification automatique des éléments, comme nous le montrons dans le chapitre 1.

---

<sup>127</sup> *HyperText Markup Langage*.



Premier exemple de dialogue (format *HTML*)

***report - dialogue***

**pil\_h\_ang\_inde**

- [mic] ENAC [oe] good morning Speedbird 5 4 4 climb for level 1 0 0 route Amboise .  
[mic]

**(no speaker)**

•

**ctrl\_h\_ang\_1**

- [mic] Speedbird euh 5 4 4 ENAC [oe] good morning euh [bb] [...] climb euh level euh  
[auto\_correction] radar identified [i] climb level 1 0 0 [b] [mic] route +[b] Lacan  
Châteaudun i call you back . [mic]

**(no speaker)**

•

**pil\_h\_ang\_inde**

- climbing for level 1 0 0 [allongement] route Amboise Lacan Châteaudun Speedbird 5  
4 4 . [mic]

**(no speaker)**

•

**ctrl\_h\_ang\_1**

- [mic] [i] euh Speedbird euh Speedbird 5 4 4 euh [...] climb level 1 4 0 euh i call you  
back for higher . [mic]

**(no speaker)**

•

**pil\_h\_ang\_inde**

- climb for level 1 4 [allongement] 0 Speedbird 5 4 4 . [mic]

\*\*\*\*\*

**Deuxième exemple de dialogue (format *HTML*)**

***report - dialogue***

**ctrl\_h\_fr\_1**

- [mic] [i] Cotam 55 42 [i] euh
- [...] [m] [...]
- transpondeur euh 31 47 [i] euh début du service de contrôle à niveau 67 0 (?)
- euh
- [i] [b] [b]
- et euh [b] joignez à Poitiers niveau euh 9 0 . [b] [mic]

**(no speaker)**

- 

**pil\_h\_fr\_inde**

- 31 47 au transpondeur et nous joignons Poitiers [pron=pi] 9: 0 Cotam 35 42 . [mic]

---

**Premier exemple de dialogue (format *trs*)**

```
<Section    type="report"    topic="to2"    startTime="798.077"
endTime="869.538">
```

```
<Turn speaker="spk15" startTime="798.077" endTime="802.568">
```

```
<Sync time="798.077"/>
```

```
<Event desc="mic" type="noise" extent="instantaneous"/>
```

```
ENAC [oe] good morning Speedbird 5 4 4 climb for level 1 0 0
route Amboise .
```

```
<Event desc="mic" type="noise" extent="instantaneous"/>
```

```
</Turn>
```

```
<Turn startTime="802.568" endTime="804.814">
```

<Sync time="802.568"/>

</Turn>

<Turn speaker="spk5" startTime="804.814" endTime="817.359">

<Sync time="804.814"/>

<Event desc="mic" type="noise" extent="instantaneous"/>

Speedbird euh 5 4 4 ENAC [oe] good morning euh

<Event desc="bb" type="noise" extent="instantaneous"/>

[...] climb euh level euh [auto\_correction] radar identified

<Event desc="i" type="noise" extent="instantaneous"/>

climb level 1 0 0

<Event desc="b" type="noise" extent="instantaneous"/>

<Event desc="mic" type="noise" extent="instantaneous"/>

route

<Event desc="b" type="noise" extent="previous"/>

Lacan Châteaudun i call you back .

<Event desc="mic" type="noise" extent="instantaneous"/>

</Turn>

<Turn startTime="817.359" endTime="818.167">

<Sync time="817.359"/>

</Turn>

<Turn speaker="spk15" startTime="818.167" endTime="821.73">

<Sync time="818.167"/>

climbing for level 1 0 0 [allongement] route Amboise Lacan  
Châteaudun Speedbird 5 4 4 .

<Event desc="mic" type="noise" extent="instantaneous"/>

</Turn>

<Turn startTime="821.73" endTime="860.092">

<Sync time="821.73"/>

</Turn>

<Turn speaker="spk5" startTime="860.092" endTime="866.724">

<Sync time="860.092"/>

<Event desc="mic" type="noise" extent="instantaneous"/>

<Event desc="i" type="noise" extent="instantaneous"/>

euh Speedbird euh Speedbird 5 4 4 euh [...] climb level 1 4  
0 euh i call you back for higher .

<Event desc="mic" type="noise" extent="instantaneous"/>

</Turn>

<Turn startTime="866.724" endTime="867.083">

<Sync time="866.724"/>

</Turn>

<Turn speaker="spk15" startTime="867.083" endTime="869.538">

<Sync time="867.083"/>

climb for level 1 4 [allongement] 0 Speedbird 5 4 4 .

<Event desc="mic" type="noise" extent="instantaneous"/>

</Turn>

</Section>

\*\*\*\*\*

### Deuxième exemple de dialogue (format *trs*)

<Section type="report" topic="to2" startTime="1112.825"  
endTime="1135.867">

```
<Turn speaker="spk7" startTime="1112.825" endTime="1130.61">
<Sync time="1112.825"/>
<Event desc="mic" type="noise" extent="instantaneous"/>
<Event desc="i" type="noise" extent="instantaneous"/>
  Cotam 55 42
<Event desc="i" type="noise" extent="instantaneous"/>
  euh
<Sync time="1115.813"/>
[... ] [m] [... ]
<Sync time="1117.394"/>
transpondeur euh 31 47
<Event desc="i" type="noise" extent="instantaneous"/>
  euh début du service de contrôle Ã niveau 67 0
<Sync time="1125.178"/>
euh
<Sync time="1126.02"/>
<Event desc="i" type="noise" extent="instantaneous"/>
<Event desc="b" type="noise" extent="instantaneous"/>
<Event desc="b" type="noise" extent="instantaneous"/>
<Sync time="1127.394"/>
et euh
<Event desc="b" type="noise" extent="instantaneous"/>
  joignez Ã Poitiers niveau euh 9 0 .
<Event desc="b" type="noise" extent="instantaneous"/>
<Event desc="mic" type="noise" extent="instantaneous"/>
</Turn>

<Turn startTime="1130.61" endTime="1131.166">
<Sync time="1130.61"/>
```

</Turn>

<Turn speaker="spk18" startTime="1131.166"  
endTime="1135.867">

<Sync time="1131.166"/>

31 47 au transpondeur et nous joignons Poitiers

<Event desc="pi" type="pronounce" extent="instantaneous"/>

9: 0 Cotam 35 42 .

<Event desc="mic" type="noise" extent="instantaneous"/>

</Turn>

</Section>



**TITLE:** *Analysis, modelling, and automated detection of disfluencies in restrained spontaneous speech dialogs: the case of Air Traffic Control*

---

**ABSTRACT:** The disfluencies are a frequently occurring phenomenon in any spontaneous speech production; they have given rise to numerous studies.

However, most of the researches about the subject relate to everyday uses of language: “small talk” dialogs, requests for schedule, speeches, etc. But what about spontaneous speech production made in a restrained framework? To our knowledge, no study has ever been carried out in this context.

Our thesis work is devoted to the linguistic and computational study of disfluencies within such a framework. These dialogs concern air traffic control, which entails both pragmatic and linguistic constraints. At first we conduct a subtle analysis of these phenomena. Then we model them to a level of abstraction, which allows us to obtain the patterns corresponding to the different configurations observed. Finally we propose a methodology for automatic processing, which is integrated into a system of automatic processing of speech, and is validated on a corpus of 400 sentences.

---

**KEYWORDS:** Spontaneous speech dialog, corpus linguistic analysis, robust understanding, Air Traffic Control, phraseology, disfluencies, language models, Natural Language Processing



**AUTEUR :** BOURAOUI Jean-Léon Mehdi

**TITRE :** *Analyse, modélisation, et détection automatique des disfluences dans le dialogue oral spontané contraint : le cas du Contrôle Aérien*

**DIRECTEUR DE THESE :** VIGOUROUX Nadine

**LIEU ET DATE DE SOUTENANCE :** Université Toulouse III, le 30 octobre 2008

---

**RESUME :** Les disfluences sont un phénomène apparaissant fréquemment dans toute production orale spontanée ; elles ont donné lieu à de nombreuses études. Cependant, la majorité des travaux de recherche sur le sujet portent sur des usages de langage quotidien : dialogues « à bâtons rompus », demandes d'horaire, discours, etc. Mais qu'en est-il des productions orales spontanées produites dans un cadre contraint ? Aucune étude n'a à notre connaissance été menée dans ce contexte.

Notre travail de thèse est consacré à l'étude linguistique et informatique des disfluences dans un tel cadre. Il s'agit de dialogues de contrôle de trafic aérien, aux contraintes pragmatiques et linguistiques. Dans un premier temps nous procédons à l'analyse fine de ces phénomènes. Ensuite, nous les modélisons à un niveau de représentation abstrait, ce qui nous permet d'obtenir les patrons correspondant aux différentes configurations observées. Enfin nous proposons une méthodologie de traitement automatique, intégrée dans un système de traitement automatique de la parole, et validée sur un corpus de 400 énoncés.

*TITRE et résumé en anglais au recto de la dernière page*

---

**MOTS-CLES :**

Dialogue oral spontané, Analyse linguistique de corpus, Compréhension robuste, Contrôle Aérien, Phraséologie, Disfluences, Modèles de langage, Traitement Automatique du Langage Naturel

---

**DISCIPLINE ADMINISTRATIVE (identique à celle de la page de titre) :**  
Informatique

---

**INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE :**

Institut de Recherche en Informatique de Toulouse (UMR 5505), Université Paul Sabatier,  
118 Route de Narbonne, 31062 Toulouse Cedex 09